

CRANFIELD UNIVERSITY

ROYAL MILITARY COLLEGE OF SCIENCE

ENGINEERING SYSTEMS DEPARTMENT

PhD THESIS

Academic Year 2002-2003

Chamnarn Pedchote

Parameter estimation for non-linear systems:

an application to vehicle dynamics

Supervisor: Dr D J Purdy

April 2003

© Cranfield University 2003. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

ABSTRACT

This work presents an investigation into the parameter estimation of suspension components and the vertical motions of wheeled vehicles from experimental data. The estimation problems considered were for suspension dampers, a single wheel station and a full vehicle.

Using conventional methods (gradient-based (GB), Downhill Simplex (DS)) and stochastic methods (Genetic Algorithm (GA) and Differential Evolution (DE)), three major problems were encountered. These were concerned with the ability and consistency of finding the global optimum solution, time consumption in the estimation process, and the difficulties in setting the algorithm's control parameters. To overcome these problems, a new technique named the discrete variable Hybrid Differential Evolution (dvHDE) method is presented.

The new dvHDE method employs an integer-encoding technique and treats all parameters involved in the same unified way as discrete variables, and embeds two mechanisms that can be used to deal with convergence difficulties and reduce the time consumed in the optimisation process. The dvHDE algorithm has been validated against the conventional GB, DS and DE techniques and was shown to be more efficient and effective in all but the simplest cases. Its robustness was demonstrated by its application to a number of vehicle related problems of increasing complexity. These include case studies involving parameter estimation using experimental data from tests on automotive dampers, a single wheel station and a full vehicle. The investigation has shown that the proposed dvHDE method, when compared to the other methods, was the best for finding the global optimum solutions in a short time. It is recommended for nonlinear vehicle suspension models and other similar systems.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Dr. D.J. Purdy, my academic advisor, for the direction, advice and supports I received from him throughout the development of this thesis. His patience and kindness have made the research a memorable experience.

I would also like to thank the Royal Thai Air Force and the Royal Military College of Science (RMCS), Cranfield University for giving me the opportunity and financial supports to undertake this research study. Thanks must also go to Mr. Philip-Toop Rose for his helps on initial experimental works and the humor we shared during the tea times.

Finally, I would like to thank my parents and my girlfriend for just being patient, tolerant and supportive.

CONTENTS

LIST OF FIGURES	v
LIST OF TABLES	ix
CHAPTER 1. INTRODUCTION	1
1.1 System Identification and Parameter Estimation	4
1.1.1 Model	4
1.1.2 A general concept review	7
1.1.2.1 A-priori knowledge	9
1.1.2.2 Input-output data	10
1.1.2.3 Model selection	11
1.1.2.4 Criterion	12
1.2 Estimator Properties	18
1.2.1 Bayes Estimator	20
1.2.2. Maximum Likelihood Estimator	21
1.2.3 Markov Estimator	23
1.2.4 Least Square Estimator	23
1.3 Parameter Estimation Methods: an review	24
1.3.1 Classical methods	25
1.3.2 Modern methods	27
1.4 Literatures Review	31
1.5 Numerical Optimisation	35
1.5.1 Choice of optimisation techniques	35
1.5.2 Motivation for the development of a new method	37
1.6 Thesis aims and Outline	38
CHAPTER 2. THE DISCRETE-VARIABLE HYBRID DIFFERENTIAL EVOLUTION	44
2.1 Introduction	45
2.1.1 Introduction to Differential Evolution	45
2.1.2 Motivations to the development of the dvHDE method	49
2.2 Representation and Initialization	55
2.3 Mutation	58
2.4 Crossover	60

2.7 Convergence of the dvHDE method	68
2.8 Numerical Examples	70
2.8.1 Single degree of freedom vehicle problem	72
2.8.2 Gear train design problem	77
2.8.3 Coil spring design problem	86
2.9 Conclusions	94
 CHAPTER 3. PARAMETER ESTIMATION OF AUTOMOTIVE DAMPERS	 99
3.1 Introduction	101
3.2 Damper Test	106
3.2.1 Experimental Set-up	106
3.2.2 Experimental Data	109
3.3 Damper Models	118
3.4 Parameter Estimation	125
3.4.1 Estimation Method	125
3.4.2 Estimation Results	128
3.4.2.1 Model quality	128
3.4.2.2 Numerical search methods: Performance comparison	136
3.5 Conclusion	154
 CHAPTER 4. PARAMETER ESTIMATION OF A SINGLE WHEEL STATION	 157
4.1 Introduction	157
4.2 A Quarter Vehicle Model	160
4.2.1 Derivation of a Quarter Vehicle Model	160
4.2.2 Influence of vehicle parameter variation to its response	172
4.2.3 Estimation of model transfer function	181
4.3 Single Wheel Station Test Rig and Instrumentation	188
4.3.1 The test rig and Instrumentation	188
4.3.2 Instrument Calibration	191
4.3.3 Measurement noise	192
4.3.4 Frequency Response	202
4.4 Estimation Method	205
4.5 Estimation Results	208

4.4 Estimation Method	205
4.5 Estimation Results	208
4.5.1 Time Domain Analysis	208
4.5.2 Estimation Results with Kalman Filter	218
4.5.3 Estimation Results: Frequency Domain	226
4.6 Conclusion	230
 CHAPTER 5. PARAMETER ESTIMATION OF A LANDROVER 110	 232
5.1 Introduction	232
5.2 Vehicle Testing	234
5.2.1 The Four Poster Test Facility	234
5.2.2 Vehicle Time Response	237
5.2.3 Vehicle Frequency Response	241
5.3 Vehicle Models	246
5.4 Parameter Estimation Method	253
5.5 Estimation Results	254
5.6 Conclusion	260
 CHAPTER 6. SUMMARY, CONCLUSIONS AND FUTURE WORK	 262
6.1 Summary	262
6.1.1 Test problems	264
6.1.2 Practical case studies	267
6.1.3 The performance of the dvHDE method	272
6.2 Conclusions	273
6.3 Future Work	275
 REFERENCES	 277
 APPENDIX A Parameter estimation of an adjustable automotive damper	 288
A1. Introduction	288
A2. Analysis of results	293
A3. Summary	297

APPENDIX B Vehicle axis systems	300
APPENDIX C Some preliminary results from a nonlinear 4DOF and 7DOF models	302
C1. Model 2: a nonlinear 4DOF model	302
C2. Model 3: a 7DOF model	304
C3. Estimation results	310
C4. Discussion of results	314

LIST OF FIGURES

Fig. 1.1 Schematic flow chart of system identification.	8
Fig. 1.2 A possible flow of information for optimisation process.	14
Fig. 1.3 A class of related estimators.	19
Fig. 2.1.1 An example of the process of generating a trial-vector.	47
Fig. 2.6.1 A flowchart of the dvHDE method.	67
Fig. 2.8.1 A single degree of freedom (1DOF) vehicle model.	72
Fig. 2.8.2 A typical plot of the best fitness values against generation number.	75
Fig. 2.8.3 A typical plot of population diversity.	75
Fig. 2.8.4 Gear train.	77
Fig. 2.8.5 Plots of two encountered convergence problems (gear train).	81
Fig. 2.8.6 Plots of results of gear train parameter estimation.	84
Fig. 2.8.7 Coil spring.	87
Fig. 2.8.8 Plots of results for coil spring problem using setting 1.	90
Fig. 2.8.9 Plots of results for coil spring problem using setting 2.	93
Fig.3.1.1 A sketch layout of typical hydraulic automotive dampers.	103
Fig. 3.1.2 Damping force-velocity characteristics of typical dampers.	105
Fig. 3.2.1 A photo of Damper Test Rig.	107
Fig. 3.2.2 Diagram of Damper Test Rig and instrumentation.	108
Fig. 3.2.3 Plots of the damper characteristics; input frequency of 1 Hz.	110
Fig. 3.2.4 Plots of the damper characteristic; input frequency of 5 Hz.	112
Fig. 3.2.5 Plots of the damper characteristic; input frequency of 10 Hz.	113
Fig. 3.2.6 Plots of the damper characteristic; input frequency of 15 Hz.	114
Fig. 3.2.7 Plots of the damper characteristic; input frequency of 20 Hz.	115
Fig. 3.2.8 Section plots of force-displacement and force- velocity.	117
Fig. 3.3.1 Simplified damper models.	122
Fig. 3.4.1 Parameter estimation technique for the damper model.	126
Fig. 3.4.2 Section plots of fitting result obtained from Model 1.	129
Fig. 3.4.3 Section plots of fitting result obtained from Model 2.	131
Fig. 3.4.4 Three-dimensional plot of fitting result obtained from Model 3.	133

Fig. 3.4.5	Section plots of fitting result obtained from Model 3	134
Fig. 3.4.6	Typical plots of MSE values and population diversity (dvHDE).	142
Fig. 3.4.7	Typical plots of MSE values and population diversity (DE method).	143
Fig. 3.4.8	Estimated parameters of Model 3 using the dvHDE method.	145
Fig. 3.4.9	Estimated parameters of Model 3 using the DE method.	146
Fig. 3.4.10	Typical plots of results by Model 5, using the DE method.	149
Fig. 3.4.11	Typical plots of results by Model 5, using the dvHDE method.	150
Fig. 4.2.1	A 7DOF vehicle model.	161
Fig. 4.2.2	A 4DOF vehicle model and dynamically equivalent model.	162
Fig. 4.2.3	A quarter vehicle model.	164
Fig. 4.2.4	A free-body diagram for the quarter vehicle model.	165
Fig. 4.2.5	Plots of system pole locations and a pair of complex poles.	169
Fig. 4.2.6	Plot of impulse response of the quarter vehicle model.	170
Fig. 4.2.7	Plot of frequency response of the quarter vehicle model.	171
Fig. 4.2.8	Influence of sprung mass variation.	174
Fig. 4.2.9	Influence of unsprung mass variation.	176
Fig. 4.2.10	Influence of spring stiffness variation.	177
Fig. 4.2.11	Influence of tyre stiffness variation.	179
Fig. 4.2.12	Influence of damping rate variation.	180
Fig. 4.2.13	Validation of the transfer function estimation method.	187
Fig. 4.2.14	Transfer function estimation; Gain and Phase errors.	187
Fig. 4.3.1	Photo of the Single Wheel Station test rig.	188
Fig. 4.3.2	Diagram of the Single Wheel Station test rig and instrumentation.	189
Fig. 4.3.3	Plots of PSD of a normally distributed sequence and its probability plot.	194
Fig. 4.3.4	Plots of PSD of sensor noise when the hydraulic pump turned off.	196
Fig. 4.3.5	Plots of PSD of sensor noise when the hydraulic pump turned on.	197
Fig. 4.3.6	Normal Probability plots, hydraulic pump turned off.	199
Fig. 4.3.7	Normal Probability plots, hydraulic pump turned on.	200
Fig. 4.3.8	Normal Probability plots, hydraulic pump turned on, filtered at 40 Hz.	201
Fig. 4.3.9	Sprung mass frequency responses.	203
Fig. 4.3.10	Unsprung mass frequency responses.	203
Fig. 4.4.1	Parameter estimation method for the Single Wheel Station model.	205

Fig. 4.5.1 Plot of time history of the sprung mass displacement (Model 1).	210
Fig. 4.5.2 Plot of time history of the sprung mass acceleration (Model 1).	210
Fig. 4.5.3 Plot of time history of the unsprung mass displacement (Model 1).	212
Fig. 4.5.4 Plot of time history of the unsprung mass acceleration (Model 1).	212
Fig. 4.5.5 Plot of time history of the dynamic tyre force (Model 1).	213
Fig. 4.5.6 Plot of time history of the sprung mass displacement (Model 3).	215
Fig. 4.5.7 Plot of time history of the sprung mass acceleration (Model 3).	215
Fig. 4.5.8 Plot of time history of the unsprung mass acceleration (Model 3).	216
Fig. 4.5.9 Plot of time history of the dynamic tyre force (Model 3).	216
Fig. 4.5.10 Plots of dynamic forces of the spring-damper subassembly.	221
Fig. 4.5.11 A typical plot of fitness values obtained from Model 3.	224
Fig. 4.5.12 A typical plot of population diversity obtained from Model 3.	225
Fig. 4.5.13 Plots of fitting result with $\text{RMS}_{\text{input}} = 7.5 \text{ mm}$.	228
Fig. 4.5.14 Plots of fitting result with $\text{RMS}_{\text{input}} = 12.5 \text{ mm}$.	229
Fig. 5.2.1 Four Poster Test Rig and a Landrover 110 front and rear suspension.	235
Fig. 5.2.2 Layout of the main equipment.	236
Fig. 5.2.3 Time history of vehicle suspension working space.	238
Fig. 5.2.4 Time history of vehicle dynamic tyre forces.	240
Fig. 5.2.5 Plots of vehicle body frequency response in bounce.	243
Fig. 5.2.6 Plots of vehicle front and rear axles frequency response in bounce.	244
Fig. 5.2.7 Plot of vehicle body (front end) frequency response.	245
Fig. 5.3.1 A four-degrees-of-freedom (4DOF) vehicle ride model.	248
Fig. 5.3.2 Free body diagram of the vehicle ride model.	250
Fig. 5.4.1 Parameter estimation method.	253
Fig. 5.5.1 Fitting result of vehicle frequency response by Model 1.	255
Fig. A1 Diagrammatic representation of the adjustable damper.	289
Fig. A2 Plot of the MSE values against the damper setting.	291
Fig. A3 Plot of the damper force against velocity.	292
Fig. A4 Plot showing the performance of the DE and dvHDE methods.	298
Fig. B1 SAE vehicle axis system.	300

Fig. C1 A four degrees of freedom model with linkages.	303
Fig. C2 A seven degrees of freedom vehicle model.	305
Fig. C3 A free body diagram for the sprung mass.	307
Fig. C4 A free body diagram for the front unsprung mass.	308
Fig. C5 A free body diagram for the rear unsprung mass.	309
Fig. C6 Fitting result of vehicle frequency response by Model 2.	311
Fig. C7 Fitting result of vehicle body frequency response by Model 3.	312
Fig. C8 Fitting result of front and rear axles frequency response by Model 3.	313

LIST OF TABLES

Table 2.8.1	Parameter estimation results for Gear train design.	79
Table 2.8.2	Available wire diameters.	87
Table 2.8.3	Optimal solutions for coil spring problem.	92
Table 3.3.1	List of damper models considered.	123
Table 3.4.1	Estimation results from the GB and DS methods.	137
Table 3.4.2	Estimation results from the DE and dvHDE methods.	138
Table 4.2.1	Representative parameters for the quarter vehicle model.	166
Table 4.3.1	The standard deviation and mean values of measurement noise.	192
Table 4.5.1	MSE values from different quarter vehicle models.	209
Table 4.5.2	Performance comparison of four different methods.	217
Table 4.5.3	MSE values based on component test data and estimated data.	220
Table 4.5.4	Estimated and component test parameter values.	222
Table 5.4.1	Performance of different numerical search methods.	258
Table A1	Estimation results by the GB and DS methods	294
Table A2	Estimation results by the DE and dvHDE methods	295

CHAPTER 1 INTRODUCTION

The optimisation of vehicle suspension systems is a time consuming and complex task. This results from the non-linear characteristics of the components involved [1,2] and the coupling together of the vehicle motions [3,4]. The use of a four-post suspension test facility helps in the initial setting-up of the suspensions, though it can still take a considerable amount of time. Considering only the pitch and bounce motions of the vehicle body, with fixed suspension springs and a 20 setting adjustable dampers [1] this would require 400 test runs to cover all possibilities. This would require a full weeks testing, which in many cases would be uneconomical. Even with all the information that would result from these tests, the selection of the best or range of good damper setting could still require a significant amount of additional processing time.

The overall objective of this work is to investigate the possibilities of reducing the required time from days to hours of testing on the four-post suspension test rig. The problem has been broken down into three stages. The first is the estimation of the vehicle parameters, from an initial vehicle test. The second is to use these parameters in a vehicle model to predict the required damper settings and possibly the suspension spring rates. The third is to repeat the testing. This would be used iteratively to rapidly achieve the best suspension settings. The iterative process is required because of the approximations built into the model used to make the predictions. This is an alternative to removing the components and testing them individually and using the

resulting data in a sophisticated model of the vehicle and suspension. The individual testing of suspension components has its own problems [5].

The results presented in this thesis work consider the initial phase of the process, the parameter estimation. Due to the nature of the problem, the models needed for the vehicle and vehicle components are nonlinear and they are built using the equations of motion for the system. In such a situation, the parameters of the models usually have a physical interpretation and accurate parameter estimation is important. An analytical solution for the estimation is difficult if not impossible to find, leading to the necessity to use numerical algorithms in searching for the model parameters.

In this work, a global optimisation algorithm named the discrete-variable Hybrid Differential Evolution (dvHDE) method is presented to solve parameter estimation problems for nonlinear systems such as the vehicle suspension system. The algorithm is based on the Differential Evolution (DE) algorithm, a relatively new method of optimisation that appeared in 1996. The proposed algorithm is fundamentally different from the original DE method in that it employs a different encoding technique to represent the solution of the parameter estimation problem considered. The dvHDE method uses an integer-encoding technique and treats the parameters involved in the same unified way whether they are integer, discrete or continuous variables, or a combination of them. The algorithm also includes two additional mechanisms to overcome possible premature convergence problems and improve convergence speed. The dvHDE method is fundamentally different from conventional methods in that it does not use gradient information of the estimation error. The unknown parameters that enter the parametric model in a nonlinear way can be estimated by the dvHDE

method, while the conventional methods may generally encounter problem of local minima and slow convergence rate. The newly developed dvHDE method has been evaluated by comparing its performance with the original DE algorithm, and a number of conventional techniques including, gradient-based and direct search methods.

In this chapter, the process of system identification and parameter estimation are introduced and finally the outline and aims of the thesis are presented. In section 1.1, some preliminary ideas of the parameter estimation task and the role it plays in system identification are discussed. The section first introduces the relevant terms relating to parameter estimation problems, for example terms such as model, black/white-box models, parameters, linear/nonlinear-in-parameter and a-priori knowledge. A brief discussion on how a model may be chosen and constructed for a particular purpose, and on important components required for solving a parameter estimation problem are also provided in this section. Section 1.2 looks at several important properties of the estimator. A commonly used class of related estimators including Bayes Estimator, Maximum Likelihood Estimator, Markov Estimator and Least Squares Estimator are also presented and deserve some comments. In section 1.3, a review of practical parameter estimation methods including classical and modern methods is provided. System identification and parameter estimation has been studied and applied to many practical problems. The interest of this thesis has been limited to wheeled vehicle dynamics, in particular the suspension system and its vertical dynamic behavior. Instead of providing a review of a large amount of literature on system identification and parameter estimation methods that have been developed over the past years, attention is only paid to methods applied to the vehicle dynamic problem. This literature review is presented in section 1.4. Section 1.5 briefly discusses choices of

different numerical optimisation methods, and lists important difficulties when applying the techniques to the problem being investigated. Finally, the aims and outline of this thesis work are then given in section 1.6.

1.1 System Identification and Parameter Estimation

1.1.1 Model

In the fields of engineering and physics, mathematical models are commonly employed as a useful and compact way for representing the information and knowledge about the system of interest. Models allow one to gain insight into the dynamic behavior of the system in order to study its behavior or to improve its operating efficiency. A model is therefore built with definite aims in mind, and its complexity is strongly dependent on its intended purposes. Broadly speaking, a system is an object in which variables of different kinds interact. A model more or less accurately describes the relationship among these variables. Whatever the structure chosen for the model, it will in general involve unknown quantities, which must be estimated from available a-priori knowledge and data. These quantities are called the *parameters* of the model, and are denoted here by a vector θ . The model structure and a particular model obtained by setting the parameters to some specific numerical value θ will be denoted by M and $M(\theta)$ respectively.

Mathematical models may be divided into different categories using specific criteria [6-9]. To choose among different model structures is a primary and important step, since it defines the choice available for the selection of the ‘best model’. This involves, for example, the selection between linear and nonlinear models, between white-box and black-box models. A model structure will be said to be linear in its inputs if its outputs satisfy the superposition principle with respect to its inputs. The principle of superposition states that if a system responds to an input u_1 with an output y_1 and to an input u_2 with an output y_2 then its response to the synthesised input $\alpha u_1 + \beta u_2$ is given by $\alpha y_1 + \beta y_2$, with α, β being constant real numbers. An outcome of the principle of superposition is the principle of homogeneity. It states that a system responds to a scaled input αu_1 with the corresponding scaled output αy_1 . In the case that either or both of the above principles fail to be true then the system is nonlinear. The modelling and the identification of linear systems is well developed and reported in many textbooks [6,10].

The construction of a mathematical model may employ the following two approaches, these being analytical or experimental, or a combination of them [10,11]. In the analytical approach, the dynamic behavior of the system is described by rules based on laws of physics (Kirchhoff’s laws, Newton’s laws, etc). Usually, the system is ‘broken’ into blocks or subsystems, whose properties are well understood from knowledge developed from past experience. The subsystems are then rejoined mathematically to obtain a model for the whole system. Such a model is called a white-box model. Another approach is the experimental approach. Basically, a model is inferred from the recorded data of the system’s input-output signals, after performing some experimentation and data analysis, respectively. The model structure

does not claim to correspond in any fundamental way to that of the system, and the parameters have no physical meaning. The construction of the model using this approach is typically called system identification, and the resultant model is a black-box model. The analytical approach should be attempted when possible, since it allows the incorporation of physical laws of nature into the model. However, for a system that is too poorly understood or too complicated to model from first principles, the system identification approach may be employed.

In the case of a structural dynamic system, the equations of motion of the studied system should be derived. The mathematical derivation is most commonly done by use of Newton's laws of motion, or energy balance methods such as Lagrange's method for a structure with complicated geometry [12-14]. Further experimentation may be performed to deduce nonlinear terms and incorporate them into the equations of motion. Furthermore, the introduction of extra variables, such as internal (unobservable) state variables, facilitates the model building when the behavior observed is very complicated.

A general form of the equations that model a real system may be expressed explicitly in term of the model parameter θ as follow.

$$\dot{\mathbf{z}}(t) = f(\mathbf{z}(t), \mathbf{u}(t), \mathbf{w}(t), \theta) \quad (1.1)$$

$$\mathbf{y}(t) = h(\mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t), \theta) \quad (1.2)$$

Where $\mathbf{z}(t)$ is the vector of states, which are the dynamic response of the model to the input vector $\mathbf{u}(t)$, and its time derivative is denoted by $\dot{\mathbf{z}}(t)$. The system states, which may be either observable or not, are determined by the structure of the function f of the input, the unobservable disturbance vector $\mathbf{w}(t)$ and the parameter vector $\boldsymbol{\theta}$. Equation (1.1) is known as the system equation. Equation (1.2) is the output equation, and is determined by the structure of the function h of the state vector, the input vector, the measurement noise $\mathbf{v}(t)$, and the parameter vector. It should be noted that the variables involved in the two equations are function of time, while the parameter vector is not. This is to emphasize that $\boldsymbol{\theta}$ is a constant parameter vector.

1.1.2 A general concept review

The aim of the system identification process is to select the best model out of a class of possible models and determine the values of the model parameters. In general, system identification consists of the following four stages:

- (a) Experimental design to collect input and output data
- (b) Model selection
- (c) Parameter estimation
- (d) Model validation

The process starts with the design of an experiment to record the input and output data. Then, a model is chosen for the system. Based on the recorded input and output data, techniques of parameter estimation are applied to estimate the parameters in the

assumed model. Finally, the model is tested to see if it is sufficient and satisfies its intended use. If it is not, the process starts again by assuming another model. In practice, system identification is iterative and a trial-and-error based procedures, which can be illustrated by Fig. 1.1. Note that a priori knowledge of the system helps in the design of the experiment, model selection and validation. Clearly, all four stages are important for system identification. The work of this thesis work focuses on the stage of parameter estimation.

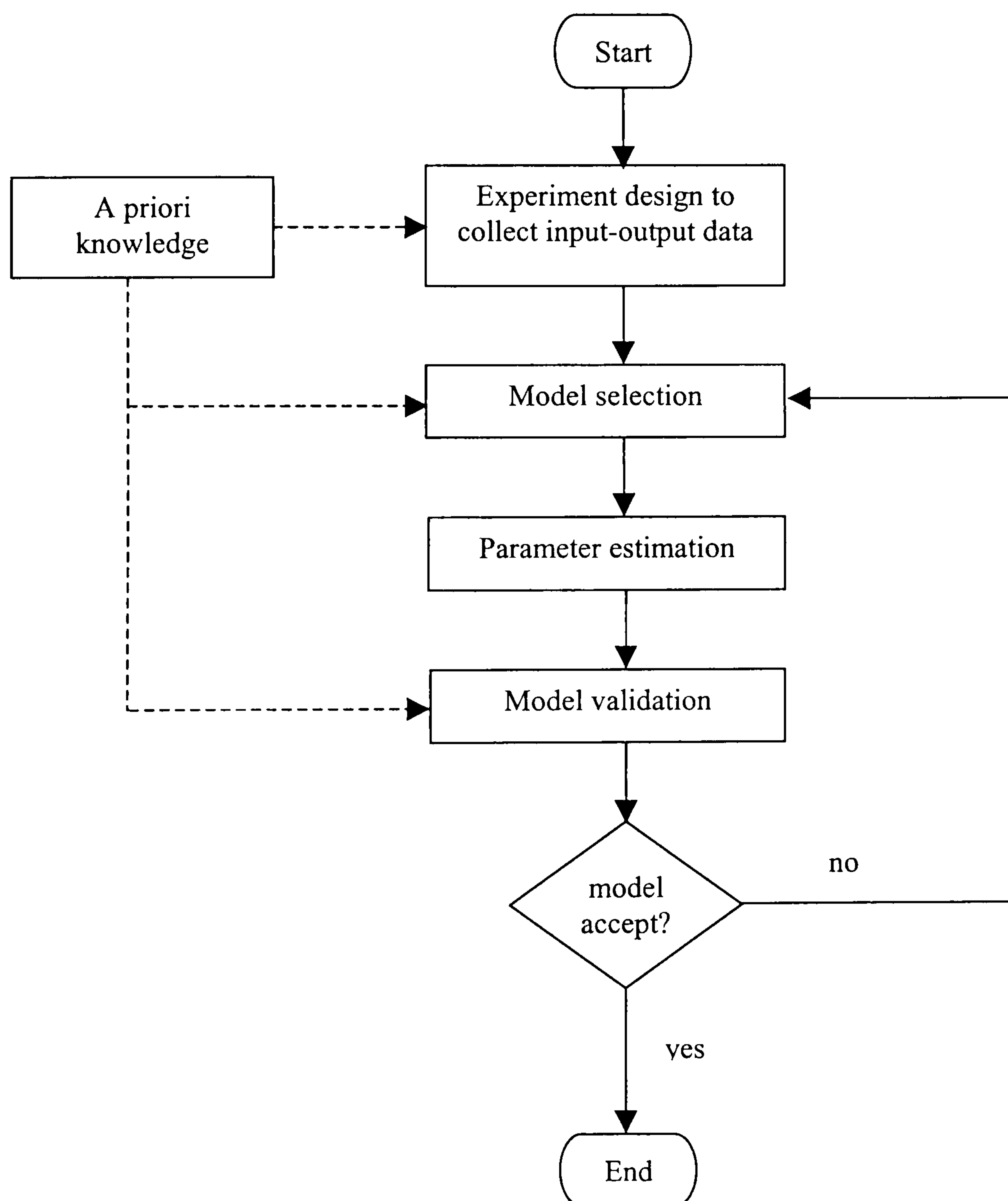


Fig. 1.1 Schematic flowchart of system identification.

A clear distinction should be born in mind between the problem of parameter estimation and system identification. When performing system identification, one has the freedom to select both the model structure and the model parameters. In other words, regarding the equation (1.1) and (1.2), one is allowed to choose the function structures for f and h in the first place and afterwards the parameter vector, θ . When the performing parameter estimation, the model is predetermined. That is the structure of the functions f and h is assumed known. The response of the model in equation (1.1) and (1.2) depends on the actual values of the components of the parameter vector. It is then the parameters that are free variables to be chosen or estimated so as to minimize the differences between the response predicted by the model and those obtained experimentally. Solving the parameter estimation problem thus requires the following:

- (a) Recorded input and output data
- (b) A chosen model
- (c) A criterion to evaluates the quality of the model.

These are discussed below.

1.1.2.1 A-priori knowledge

Before any steps in the system identification and parameter estimation process is carried out, a collection of all available a-priori knowledge about the system being studied is crucial. The most important one is the acknowledgement of the intended operations the identified model needs to perform. A-priori knowledge provides a tentative model structure, and some ideas about the experiment and input

requirements. Any previous understanding about the system is regarded as a-priori knowledge. It normally includes information such as system operating frequency range, its natural frequencies, the number of active modes, the linearity, the stationarity, and statistics of the noise perturbed over the system. An expectation of the kind of error that will occur during measurement, estimation, optimization, and numerical errors can also be considered as a-priori knowledge.

1.1.2.2 Input-output data

The input-output data is obtained after performing some experimentation on the system. The design of the experiment is important since a substantial amount of information used in the identification/parameter estimation process is extracted from the data obtained. The experimentation should be designed in such a way as to allow as much information as possible to be portrayed in the data, taking into account the limits imposed by the constraints on the allowable experimental conditions [9,15,16]. This includes making decision on such issues as what type of input signal should be used, where the measurement should be made, and how the signals should be filtered and sampled. The input characteristics are of primary importance because they dictate the quality of the results of an identification/parameter estimation method. A complete optimization of input signal should involve both time and frequency domain behavior [8]. Generally, the input should be able to persistently excite all the relevant dynamics of the system. Isermann [17], Leontaritis and Billings [18] suggested that a signal is used to persistently excite a system if its power spectrum density does not vanish for the system operating frequency range. A further detailed discussion on optimal input and experiment design can be found in [6-9, 16-18].

1.1.2.3 Model selection

Before arriving at one particular model, a primary and very important step is to choose the class of models to be considered. This should be guided by the intended use of the model. If the primary task of the model is to perform an accurate prediction of a future outcome of the system based on some past observations an input-output type of model, also known as black-box model, can be used. In this case, a model is proposed, which allows a sufficient description of the observed input and output measurements. The model parameters do not conveying any physical meaning and as such their values are not important. As a consequence, this type of model offers flexibility in choosing the model structures such that the models are linear-in-parameters, and therefore the parameter estimation techniques for linear systems, which have been well studied [19-24], can be applied. Methods such as the Least Mean Square (LMS) algorithm and Recursive Least Square (RLS) are typical examples. They are on-line approaches, in which explicit formulas can be derived for the parameter θ . The parameter estimation problem becomes much less complex when the model is linear-in-parameter. If a nonlinear system is modeled such that the parameters enter linearly, then techniques similar to those for linear systems can be used to estimate the model parameters. A model is said to be linear-in-parameter, if there is a linear relation between the model parameters and the model output. These types of models are usually found in adaptive controllers, as their important task is only to give an accurate control action [25,26].

On the other hand when the task of the model is to provide a better understanding of the physics of a system, then a structural model, also called a white-box model, is needed for which each parameter and term has a physical interpretation. Therefore,

accurate parameter estimates are important. In engineering dynamics for example these models are usually useful in the field of fault detection since any fault can be associated with a parameter modification [27]. It is often the case that when a structural model is employed in modeling for these purposes it is not possible to have the model linear-in-parameter, and methods similar to those for linear systems may not be appropriate and/or sufficient. The method such as Least Square (LS) and Maximum Likelihood estimators can be used. When the model is nonlinear-in-parameter, it is difficult if not impossible to find an analytical solution for the parameter estimation problem. As a consequence, a numerical search algorithm has to be used in searching for the best model parameters. The gradient-based approach such as Gauss-Newton method and its variants [19,28] can suffer from being trapped at local minima of the search spaces. There has not been much research undertaken on solving the parameter estimation problem for nonlinear systems modeled by parametric models in which the parameters do not enter linearly. In this work, a global optimization algorithm called the Differential Evolution (DE) has been modified to estimate the parameters of this kind of nonlinear system.

1.1.2.4 A criterion to evaluate the model quality

After the class of the model has been determined, the next step is to choose the best model in the class thus defined. This raises the question of the definition of the measure or criterion required to compare the performance of the competing models. The criterion is usually in the form of a scalar function J of the parameters, called the cost function. Assume that the cost function is to be minimized, $M(\theta_1)$ is then better than $M(\theta_2)$ in the sense of the criterion associated with J if $J(M(\theta_1)) < J(M(\theta_2))$.

There are two issues regarding the selection of a criterion needing to be addressed. One is the properties of the estimator required for the parameter estimation, this will be further discussed in section 1.2. The other is concern about the method, that is how the parameter estimation procedure is carried out and this will be discussed in section 1.3.

The optimal value of θ will certainly depend on the criterion chosen. Various approaches can be used to formulate the criteria depending on the available a priori information of the system being studied. Different choices of the criteria lead to different estimation methods. Over the years several methods have been developed, the ones extensively used however are brief mentioned here.

Considering the situation where the model is of parallel type as depicted in Fig. 1.2. By ‘parallel’ here means that the model is subjected to the same input $u(t)$ as the system, and the initial conditions are taken to be zero as the system is assumed to be stable enough for their transient effect to be neglected. The model output is obtained according to equation (1.1) and (1.2), and it is denoted by $\hat{y}(t | \theta)$ to emphasis that at time t the model represented by $M(\theta)$ will generate a prediction $\hat{y}(t | \theta)$ based on the given parameter vector θ . To assess the goodness of this prediction its difference with the system output $y(t)$ is used. Hence the error or cost can be defined as follows.

$$e(t | \theta) = y(t) - \hat{y}(t | \theta) \quad (1.3)$$

This is called the prediction error $e(t | \theta)$ at time t , and is again written explicitly as a function of the parameter vector, θ . This means $M(\theta_1)$ and $M(\theta_2)$ will produce two

different sequences of the prediction errors $e(t, \theta_1)$ and $e(t, \theta_2)$ respectively. For models that this error does not depend on previous output values is called an output error [6,7]. The error defined in equation (1.3) can be used as a comparison means between different parameters. A selection of the parameter for which the sequence of the output errors becomes as small as possible is the aim of a particular parameter estimation methods. Generally, from a very large set of the parameter vector (i.e. from θ_i where $i = 1, 2, 3, \dots$), one searches for the parameter vector that produce the sequence with the smallest error.

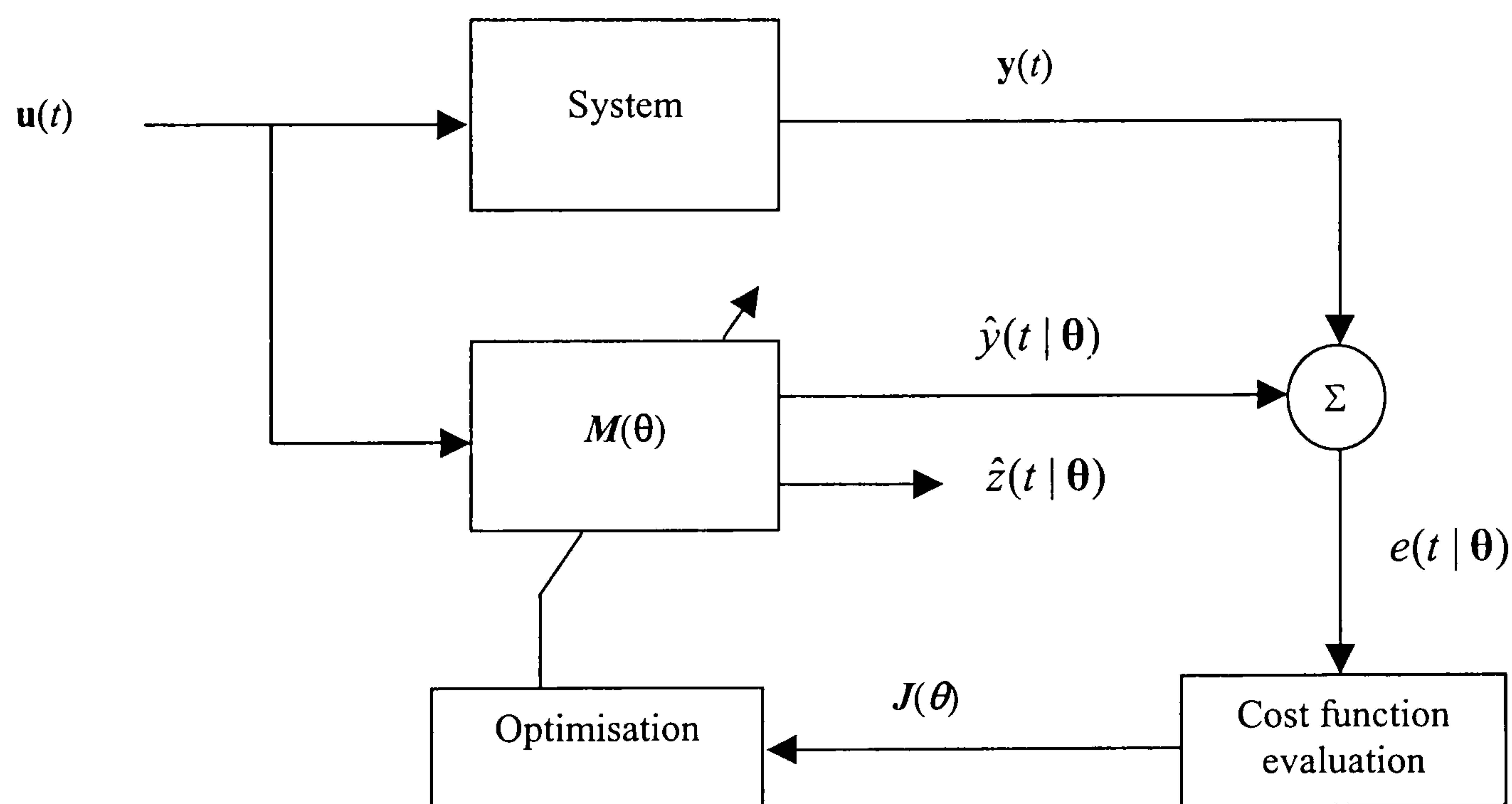


Fig. 1.2 Possible flow of information for optimisation

The Least Square method finds the smallest possible sequence of errors by trying to minimize the following scalar function

$$J_{LS} = \frac{1}{N} \sum_{i=1}^N (y(i) - \hat{y}(i | \theta))^2 \quad (1.4)$$

where N is the number of data points. By taking the derivatives of equation (1.4) with respect to the parameters and equating them to zero one arrives at a system of N least square equations with n_j unknowns (n_j being the number of parameters). Its solution gives the required parameter values [6-9].

Suppose that the output (prediction) error has a conditional probability density function depending on the time t , the parameter vector θ , and the input. In addition, assuming that the prediction errors for two different time instants are independent and that the probability density function (PDF) for an individual instant is given by an arbitrary function P , the joint probability density function for all the error values is given by

$$\text{JPDF} = \prod_{i=1}^N P(y(i) - \hat{y}(i | \theta)) \quad (1.5)$$

The value of this function gives the probability that the output $\hat{y}(t | \theta)$ of the model at the instant i is the output of the system. If the reference is to the parameter vector θ , in the equation (1.5) is known as the likelihood of the parameter vector to give a model that replicates the system output. The parameter estimate $\hat{\theta}$ that maximizes equation (1.5) is called the maximum likelihood estimate [7,10].

Finally, the error sequence $e(t)$ can be made small when a parameter estimate $\hat{\theta}$ gives a sequence that is not correlated with another sequence $\varsigma(t)$ containing information relevant to the identified system. Generally this is formulated as

$$\frac{1}{N} \sum_{i=1}^N e(t) \varsigma^T(t) = 0 \quad (1.6)$$

The solution of equation (1.6) with respect to θ gives the required parameter estimate $\hat{\theta}$. Intuitively this means that the predictors $\hat{y}(t | \theta)$ utilize all the information at time t , rendering the prediction error independent of the system information contained in $\varsigma(t)$. The sequence $\varsigma(t)$ contains the instruments to obtain $\hat{\theta}$ and the method is known as the instrumental variable method [6,7,16].

The methods described so far work well when the system is linear-in-parameter and analytic. This is because they require the calculation of the derivatives of the objective function with respect to the parameters. In such cases the complexity only increases with the number of parameters to be identified, as the linearity ensures quadratic objective functions and continuity ensures existence of the derivatives.

When the model is nonlinear-in-parameter, solving for the parameter estimation problem is usually performed iteratively using a numerical optimisation algorithm. In this case, the criterion based on the mean square error can be used, though many others are possible [6-10]. The mean square error assumes that an ideal predictor $\hat{y}_m(t | \theta)$ of the observed sequence $y(t)$, the one that minimizes the expected value in equation (1.7) of their square difference [15].

$$\text{MSE} = \mathbf{E}((y(t) - \hat{y}(t | \theta))^2) \quad (1.7)$$

where \mathbf{E} denotes the expectation operator. For computational purposes equation (1.7) can be discretised and normalized in the following way.

$$\text{MSE} = \frac{100}{N\sigma_y^2} \sum_{i=1}^N (y(i) - \hat{y}(i | \boldsymbol{\theta}))^2 \quad (1.8)$$

where σ_y^2 is the variance of the measured output and N the number of measured points.

After obtaining the optimal parameter estimated for a chosen model, the next step is check whether the performance of the model satisfies the objectives set. Verification of the quality of fit and adequacy of the model can be made, first by checking the residual, which is the difference between the measurement responses and the modeled outputs. The residual may possess some known statistical properties that may indicate whether there still exist modeling-errors for the given input signals. This information can then be fed back into the model building process. The next check is to examine the performance of the model with a new set of measurements, bearing in mind that the new set of measurement is obtained from the same experimental conditions. The results of the identification/parameter estimation process, finally, can be compared with parameter values found using other measuring techniques, if possible.

1.2 Estimator properties

Before looking at the estimation methods, this sub-section provides a brief discussion on several important properties of estimators required for the parameter estimation problems. It is certainly true that an estimator should extract and use all information contained in the measurements. It is also desirable that the average of the parameter estimated converges to their true values, that is to say that the estimator is unbiased. In general, the expected value of an estimator should converge to the true value if the number of measurements increases to infinity. This may be interpreted by the mathematical representation below.

$$\mathbf{E}\{\boldsymbol{\theta}_m\} = \boldsymbol{\theta}^* \quad (1.9)$$

where \mathbf{E} denotes the expectation operator, $\boldsymbol{\theta}^*$ is the exact value of the parameters, and $\boldsymbol{\theta}_m$ is the estimated values obtained from m measurements. An estimator that offers this property is called an asymptotically unbiased estimator. An important property of the estimator is consistency. The consistency of an estimator may be defined, for a given estimator, by

$$\lim_{m \rightarrow \infty} \mathbf{P}[|\boldsymbol{\theta}_m - \boldsymbol{\theta}^*| > \varepsilon] = 0 \quad \text{for } \varepsilon > 0 \quad (1.10)$$

where \mathbf{P} denotes the probability operator. In practice, it is not only important to have small modelling error, but also to have small uncertainties on the estimated parameters. Therefore an efficient estimator that produces small uncertainty is required. This can be better explained by considering two unbiased estimators, $\boldsymbol{\theta}_1$ and

θ_2 with mean value θ and covariance matrices C_{θ_1} and C_{θ_2} . θ_1 is an efficient estimator if the eigen-value of the matrix C_{θ_2} is greater than or equal to the eigen-value of C_{θ_1} , or mathematically expressed as $|C_{\theta_2} - C_{\theta_1}| \leq 0$ for any estimator θ_2 with

$$C_{\theta_1} = E\{(\theta_1 - \theta)(\theta_1 - \theta)^T | \theta\} \quad (1.11a)$$

$$C_{\theta_2} = E\{(\theta_2 - \theta)(\theta_2 - \theta)^T | \theta\} \quad (1.11b)$$

Another important property of the estimator is robustness. That is (some of) the estimator properties are still valid even when the assumptions made in its construction are no longer applicable. If it can be proved that an estimator is robust with respect to consistency, then we can be sure that it will converge to the true value as the number of measurement increases.

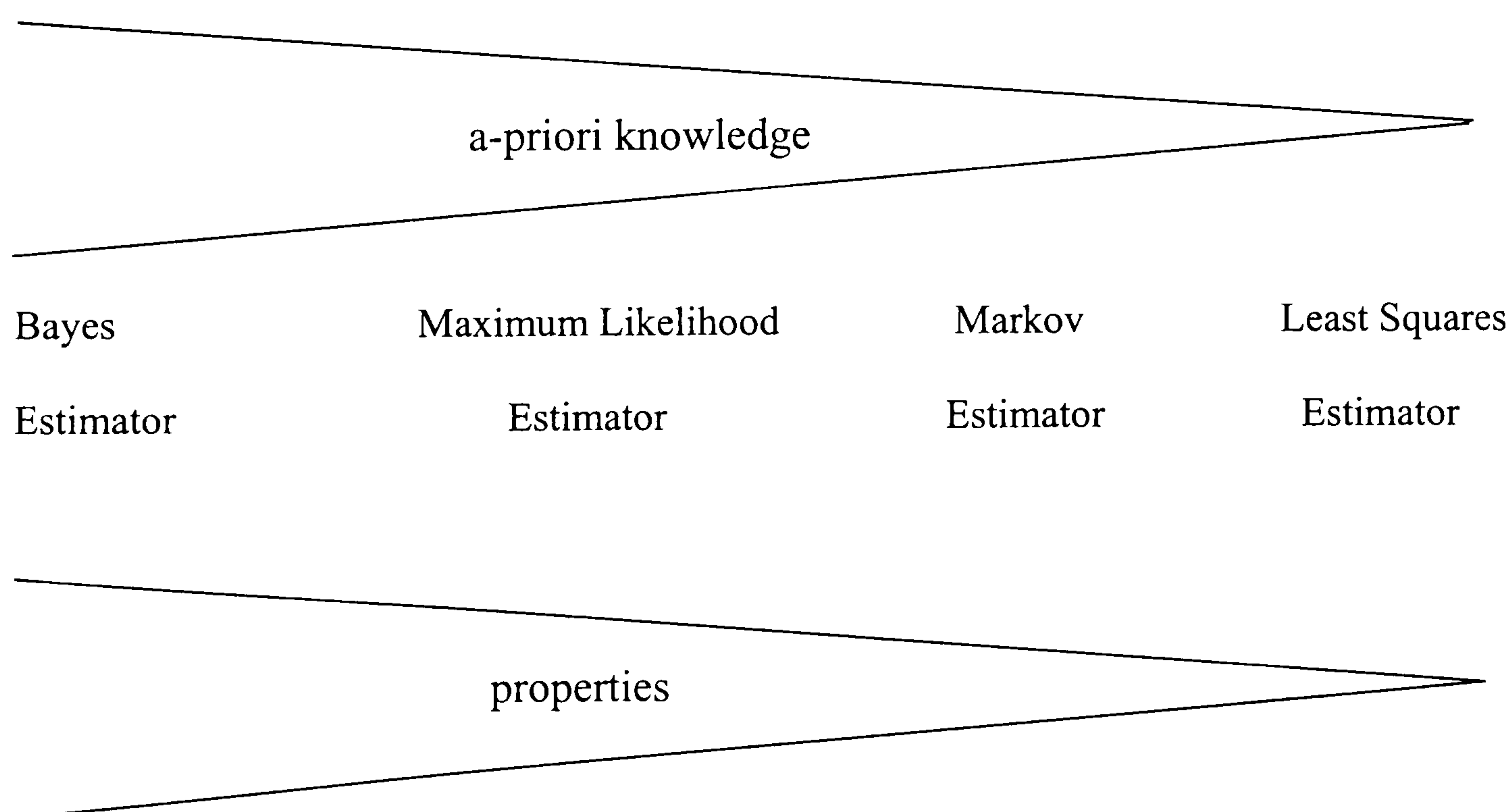


Fig. 1.3 A class of related estimators (taken from [6])

Fig. 1.3 shows a commonly used class of related estimators, which offers nearly all the properties of an estimator required. At one extreme is the Bayes Estimator, which requires the most a-priori information, at the other is the Least Squares Estimator, which is the only one in the class that can be used if there is no a-priori information available.

1.2.1 Bayes Estimator

The bayes Estimator combines a priori knowledge of the parameters with information from the measurements. To use the Bayes estimator, both the a priori probability density function (p.d.f) of the parameters and the p.d.f of the noise on the measurements are required. To determine the estimated parameters, θ_e , it is necessary to form an objective criterion, for example the minimisation of a cost function $J(\theta_e | \theta^*)$ which describes the cost of selecting the parameters θ_e if θ^* are the true but unknown parameters. Denoting probability of a parameter given a set of measurement, y_m , as

$$P[\text{parameter} | \text{measurement}] = P[\theta_e | y_m]$$

The minimisation is carried out with $P[\theta_e | y_m]$ as a weighting function

$$\theta_e = \min_{\theta} \int_{\mathbf{R}} J(\mathbf{r} | \theta^*) P[\mathbf{r} | y_m] d\mathbf{r} \quad (1.12)$$

Fries and Cooperrider [29] employed a Bayes estimator to rail vehicle problem. In many problems the required a priori information for Bayes Estimator is unavailable,

and this is one of the main reasons why Bayes estimators are rarely used in practice [30].

1.2.2 Maximum Likelihood Estimator

The Maximum Likelihood Estimator (MLE) requires knowledge of only the p.d.f of the measurement noise, with the assumption that the p.d.f of the parameter is a uniform distribution. From Bayes rule:

$$\mathbf{P}[\boldsymbol{\theta} | \mathbf{y}_m] = \frac{\mathbf{P}[\mathbf{y}_m | \boldsymbol{\theta}] \mathbf{P}[\boldsymbol{\theta}]}{\mathbf{P}[\mathbf{y}_m]} \quad (1.13)$$

and by assuming that the parameter is uniformly distributed, then equation (1.7) becomes

$$\mathbf{P}[\boldsymbol{\theta} | \mathbf{y}_m] = k_p \mathbf{P}[\mathbf{y}_m | \boldsymbol{\theta}] \quad (1.14)$$

where k_p and $\mathbf{P}[\boldsymbol{\theta}]$ are constants and $\mathbf{P}[\mathbf{y}_m]$ is seen as a scale factor.

The function $\mathbf{L}[\mathbf{y}_m | \boldsymbol{\theta}] = \mathbf{P}[\mathbf{y}_m | \boldsymbol{\theta}]$ is called the likelihood function. The maximum likelihood estimate $\boldsymbol{\theta}_{ML}$ of $\boldsymbol{\theta}$ is given by the value of $\boldsymbol{\theta}$ which maximises the likelihood function $\mathbf{L}[\mathbf{y}_m | \boldsymbol{\theta}]$. In most problems, the likelihood function is replaced by $-\ln(\mathbf{L}[\mathbf{y}_m | \boldsymbol{\theta}])$ which is known as negative of log-likelihood function. This is applicable because the natural logarithm is monotonic increasing function so minimisation of $-\ln(\mathbf{L}[\mathbf{y}_m | \boldsymbol{\theta}])$ is the same as maximisation of $\mathbf{L}[\mathbf{y}_m | \boldsymbol{\theta}]$. Many of the

MLE properties have been proven under conditions of independent, identically distributed noise on the measurements, and of a log-likelihood function, which is differentiable twice [6,30,31]. The MLE can be proven to be unique under the conditions described. A MLE is an asymptotically unbiased, consistent and efficient estimator [3]. This means that

$$\lim_{m \rightarrow \infty} \mathbf{P}[|\theta_m - \theta| > \varepsilon] = 0 \quad \forall \varepsilon > 0 \quad (1.15)$$

and the covariance matrix of a MLE converges asymptotically, as the number of measurement tends to infinity, to

$$\mathbf{C}_\theta = \text{inv}(\mathbf{F}_i) \quad (1.16)$$

where $\text{inv}(\mathbf{F}_i)$ is the inverse of the so-called Fisher information matrix. The matrix \mathbf{F}_i is a measure of the amount of information presented in the measurements in relation to the parameters. It is important to note that it is trivial to generate a biased estimator with a zero covariance matrix. However it is impossible for an unbiased estimator to have a covariance matrix smaller than the inverse of the Fisher information matrix. This means that there is a lower bound on the covariance matrix for a given set of measurements. This is called the Cramer-Rao lower bound. Covariance of the MLE approaches Cramer-Rao lower bound asymptotically, which means that the MLE is asymptotically efficient. MLE has become very popular because of the attractive properties it possesses [7,8,32].

1.2.3 Markov Estimator

The Markov estimator, which is derived from the MLE by further placing assumption on measurement noise to be additive and be characterised by a normal distribution with an a priori known mean value and standard deviation, can be used if the covariance matrix of the noise is known. The covariance matrix in a Markov estimator is seen as a weighting matrix when forming the cost function. If the weighting matrix is not the covariance matrix but an arbitrary positive definite matrix, the estimator is known as a weighted least square estimator (WLS). If measurement noise is additive and normally distributed with zero mean and a known covariance, then the Markov estimator and MLE are identical. In practice, the Markov estimator is also used in situations where the noise properties are only partially known [8]. It will not then be certain that the properties of the MLE are valid for the Markov estimator; the estimates can be biased, inconsistent and inefficient.

1.2.4 Least Square Estimator

The Least Square Estimator (LSE), which is the simplest and most commonly used, is equal to the MLE in the special case of Gaussian white noise on the measurements. The properties of the LSE can be guaranteed only in this situation, and in any other they should be investigated explicitly [7,8]. In the commonly occurring case of independent normally distributed noise on the measurements, the MLE reduces to a WLS. However, even if the situation does not match this assumption, the LSE has some advantageous properties. If the WLS is properly applied it will still be consistent

even when the disturbing noise is other than normally distributed, but it is no longer asymptotically efficient, and consequently the uncertainty on the estimate will increase compared with that of the corresponding MLE. LSE offers a much simpler implementation in practical use, however LSE has some severe drawbacks when compared to other cost functions, and the numerical advantages become progressively less important with the appearance of more and more powerful optimisation routines in scientific libraries, which permit the solution of more complex minimisation problems [33-35]. The most severe drawback of LSE compared to the least absolute values estimator is its sensitivity to gross measurement error, normally called the outliers. If a small fraction of observations contain large errors, significant deviations can be introduced in the LSE.

1.3 Parameter Estimation Methods: an review

Assuming that the chosen model structure is sufficient to represent the dynamics of the system of interest, the task is considered here to obtain a method for identifying the model parameters, which best describe those dynamics. Parameter estimation methods may be classified into two groups. The first group is the classical techniques, which include Least Squares, Weighted Least Squares, Equation Error and Output Error methods. The second group is the modern techniques, which include the Maximum Likelihood method and the Extended Kalman filter.

1.3.1 Classical methods

The Least Square method, which is mainly known in its application to the curve fitting and regression analysis, is simple and perhaps the first approach to the concept of optimality, see for example [30]. After experimentation has been conducted to collect the system input and dynamic responses, those measured signals can be used to obtain the model outputs and then select the parameters. This can be done in a number of ways, either directly, using the recorded time histories such as in [36], or after transformation of the data into the frequency domain, as in [29].

For generality, referring again to the system equation (1.1) and output equation (1.2);

$$\dot{\mathbf{z}}(t) = f(\mathbf{z}(t), \mathbf{u}(t), \mathbf{w}(t), \boldsymbol{\theta})$$

$$\mathbf{y}(t) = h(\mathbf{z}(t), \mathbf{u}(t), \mathbf{v}(t), \boldsymbol{\theta})$$

The Equation Error method assumes that, apart from the unknown model parameters $\boldsymbol{\theta}$, it is possible to obtain all system states $\mathbf{z}(t)$ and its time derivatives $\dot{\mathbf{z}}(t)$, and the input $\mathbf{u}(t)$. The criterion for the optimisation is formulated as

$$J(\boldsymbol{\theta}) = \sum_{k=1}^N \mathbf{e}_k^T(\boldsymbol{\theta}) \mathbf{W}_k \mathbf{e}_k(\boldsymbol{\theta}) \quad (1.17)$$

with

$$\mathbf{e}_k(\boldsymbol{\theta}) = \dot{\mathbf{z}}_k - f(\mathbf{z}_k, \mathbf{u}_k, \boldsymbol{\theta}) \quad (1.18)$$

where \mathbf{W}_k is a weighting matrix, the subscript $k = 1, 2, 3, \dots, N$ denotes the discrete-time version of the variables, and N denotes number of data points. Equation Error method can account for modeling error and the measurement noise. Its advantages are due to a computational simplicity and it is effective in the presence of process noise. The method was employed by Lin and Kortum [36] to obtain estimates of suspension parameters. There are some disadvantages for this method, apart from the measured input $\mathbf{u}(t)$, the measurements of the state variables $\mathbf{z}(t)$ and its derivatives $\dot{\mathbf{z}}(t)$ are also required. According to equation (1.18), the method is thus biased by the measurement noise of the variable involved. To reduce the effect, it therefore requires accurate measurements of the derivatives $\dot{\mathbf{z}}(t)$, state variables $\mathbf{z}(t)$, and input $\mathbf{u}(t)$. If some of $\dot{\mathbf{z}}(t)$ or $\mathbf{z}(t)$ are not directly available or expensive to be measured, these variables need to be estimated. The method of obtaining the variables must provide good results in order to obtain an accurate parameter values [37].

For the Output Error method, instead of requiring all the measurements of the system states and derivatives, the cost function is formulated using the estimated model outputs and the corresponding system measured responses. The model outputs are obtained using the output equation (1.2) with the measured input and the estimated model parameters $\boldsymbol{\theta}$. The cost function is formed in a similar way as in equation (1.17), but with

$$\mathbf{e}_k(\boldsymbol{\theta}) = \mathbf{y}_k - \hat{\mathbf{y}}_k(\boldsymbol{\theta}) \quad (1.19)$$

where \mathbf{y}_k is the measurement at a discrete time k and $\hat{\mathbf{y}}_k$ is the predicted or estimated measurement using the proposed model and the best guess of $\boldsymbol{\theta}$ currently available. The best value of $\boldsymbol{\theta}$ is then chosen to be the one that minimizes the cost function. The measurement requirements for the Output Error method are greatly relaxed over the Equation Error method. The measurement noise contaminates the system outputs rather than the states, and the derivatives. The method works well in the presence of the measurement noise, however it will give a biased estimate in the presence of process noise including unmeasured inputs, unmodeled system dynamics, and errors in input measurements [7,37].

1.3.2 Modern methods

To deal with measurement and process noises together, the concept of the Kalman Filter will be investigated and briefly presented here. Detailed treatment of the Kalman Filter can be found in the literature, for example [38-40]. The Kalman Filter has been successfully applied to many engineering problems, the majority of which are as a state estimator or state observer. In [41,42] a Kalman Filter was implemented as part of vehicle control and dynamics problems.

Assuming the process and measurement noise are additive, rewrite the system and output equations as follows,

$$\dot{\mathbf{z}}(t) = \mathbf{f}(\mathbf{z}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \mathbf{G} \mathbf{w}(t) \quad (1.20a)$$

$$\mathbf{y}(t) = \mathbf{h}(\mathbf{z}(t), \mathbf{u}(t), \boldsymbol{\theta}) + \mathbf{v}(t) \quad (1.20b)$$

If the system behavior is assumed to be linear, equations (1.20a) and (1.20b) reduce to

$$\dot{\mathbf{z}}(t) = \mathbf{A} \mathbf{z}(t) + \mathbf{B} \mathbf{u}(t) + \mathbf{G} \mathbf{w}(t) \quad (1.21a)$$

$$\mathbf{y}(t) = \mathbf{C} \mathbf{z}(t) + \mathbf{D} \mathbf{u}(t) + \mathbf{v}(t) \quad (1.21b)$$

where \mathbf{A} and \mathbf{B} are system and input matrices respectively, $\mathbf{w}(t)$ is process noise, assumed to be Gaussian white noise, matrix \mathbf{G} , which acts as a filter, takes into account the case when the process noise is not white. \mathbf{C} and \mathbf{D} are the output and direct transmission matrices respectively, and $\mathbf{v}(t)$ accounts for the measurement noise, assumed to be Gaussian white noise. Thus, we have

$$\mathbf{E} \left\{ \begin{bmatrix} \mathbf{w} \\ \mathbf{v} \end{bmatrix} \begin{bmatrix} \mathbf{w} & \mathbf{v} \end{bmatrix}^T \right\} = \begin{bmatrix} \mathbf{Q} & 0 \\ 0 & \mathbf{R} \end{bmatrix} \quad (1.22)$$

i.e $\mathbf{w}(t)$ and $\mathbf{v}(t)$ are white noise and assumed uncorrelated with each other. The estimation problem is approached by formulating the cost function similar to equation (1.17), but with

$$\mathbf{e}(\boldsymbol{\theta}) = \mathbf{y}_k - \hat{\mathbf{y}}_k(\boldsymbol{\theta}, \mathbf{y}_k) \quad (1.23)$$

It is interesting to note that the predicted output $\hat{\mathbf{y}}_k$ is computed using both the proposed model and the measured output, \mathbf{y}_k . Given measurement \mathbf{y}_k , the Kalman

Filter gives an answer to the question 'what will be the best estimate' of the states $\mathbf{z}(t)$ by letting the estimate of $\mathbf{z}(t)$, $\hat{\mathbf{z}}(t)$, be determined from

$$\dot{\hat{\mathbf{z}}}(t) = \mathbf{A}\hat{\mathbf{z}}(t) + \mathbf{B}\mathbf{u}(t) + \mathbf{K}(\mathbf{y}(t) - \hat{\mathbf{y}}(t)) \quad (1.24a)$$

$$\hat{\mathbf{y}}(t) = \mathbf{C}\hat{\mathbf{z}}(t) + \mathbf{D}\mathbf{u}(t) \quad (1.24b)$$

where \mathbf{K} , is called the Kalman gain, is chosen such that the 'size' of the covariance of the state error in equation (1.25) is minimal.

$$\mathbf{C}_{KF} = \mathbf{E}\{[\tilde{\mathbf{z}}(t) \tilde{\mathbf{z}}(t)^T]\} \quad (1.25)$$

with

$$\tilde{\mathbf{z}}(t) = \hat{\mathbf{z}}(t) - \mathbf{z}(t) \quad (1.26)$$

It should be born in mind that the system matrix \mathbf{A} is certainly dependent on the system parameters, $\boldsymbol{\theta}$. The input matrix \mathbf{B} and output and transmission matrices, \mathbf{C} and \mathbf{D} , could also be a function of system parameters. This means, apart from depending on measurement and process noise the variance matrices, \mathbf{Q} and \mathbf{R} , the characteristic of the Kalman Filter is also dependent on the assumed matrices \mathbf{A} , \mathbf{B} , \mathbf{C} and \mathbf{D} . To apply the Kalman Filter to problems of parameter estimation, the parameter vector is augmented as state variables. This leads to the extended version of Kalman Filter, called the Extended Kalman Filter (EKF). In a simple, linear problem when the system parameters can be assumed constant, the estimation procedure can be considered as a special case of the general state estimation, where the parameters are a set of random variable satisfying the differential equation $\dot{\boldsymbol{\theta}}(t) = 0$. To account for the

time varying parameters, unknown parameter may be modeled in the form $\dot{\boldsymbol{\theta}}(t) = \mathbf{w}_{\boldsymbol{\theta}}(t)$, where $\mathbf{w}_{\boldsymbol{\theta}}(t)$ is random process with Gaussian properties.

By combining the parameter vector and state vector in a composite state vector, we then obtain the continuous-discrete model as follows

$$\dot{\mathbf{z}}^{com}(t) = f(\mathbf{z}(t), \boldsymbol{\theta}) + \mathbf{w}^{com}(t) \quad (1.27)$$

$$\mathbf{y}_k = [\mathbf{I} \quad \mathbf{C}_k(\boldsymbol{\theta})] \mathbf{z}_k^{com} + \mathbf{v}_k \quad (1.28)$$

where

$$\mathbf{z}^{com}(t) = \begin{bmatrix} \boldsymbol{\theta}(t) \\ \mathbf{z}(t) \end{bmatrix} \quad (1.29)$$

$$f(\mathbf{z}(t), \boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0} \\ \mathbf{A}(\boldsymbol{\theta})\mathbf{z}(t) + \mathbf{B}(\boldsymbol{\theta})\mathbf{u}(t) \end{bmatrix} \quad (1.30)$$

$$\mathbf{w}^{com}(t) = \begin{bmatrix} \mathbf{w}_{\boldsymbol{\theta}}(t) \\ \mathbf{0} \end{bmatrix} \quad (1.31)$$

As can be seen, the estimation problem is a non-linear, even if we are considering a linear system, because the product $\mathbf{A}(\boldsymbol{\theta})\mathbf{z}(t)$ is a non-linear function of $\boldsymbol{\theta}$ and $\mathbf{z}(t)$.

The EKF approximates the non-linear term $f(\mathbf{z}(t), \boldsymbol{\theta})$ by a Taylor series expansion about the current known state vector $\mathbf{z}^{com}(t)$. The order of the expansion chosen is dependent on the nature of the problem being dealt with. One derivation of the EKF giving more detail can be found in [38]. The work by Baguley [43] provides some

aspect of the EKF algorithm to be implemented in practice. Although the EKF method is applicable to fully non-linear systems [44], and is straightforward from the computational viewpoint, however, the EKF is an approximate technique and there is no guarantee that the estimated parameters will be close to the true values. The method requires a-priori covariance for the parameters, which are normally unknown and also 'good' a-priori values of the parameter themselves in order for the algorithm to avoid poor convergence or failure [29,42,45-46].

1.4 Literature Review

System identification has a broad field of applications including system dynamics, structural mechanics and acoustics. Mathematicians and engineers have developed a number of various approaches to address the identification problem. The identification of a linear time-invariant system is well understood and theoretically well developed. It is however not true for the identification of a nonlinear system. Rather than giving an exhaustive survey on identification techniques developed in those fields, this section is focus at presenting a brief review of the techniques used in the field of vehicle dynamics. Many techniques have been developed over the years, however, they may be distinguished into three major applications, as follows.

The first application is non-parametric analysis of measured vehicle signals. The non-parametric analysis usually results in graphical representations that may provide some knowledge of the vehicle behaviour on its suspension without the need of an explicitly

formulated mathematical model. However, these non-parametric techniques are usually not capable of sufficiently describing the system for the solution of design, optimisation and control problems. The techniques are broadly used in industry and are well known in the literature [47-49].

The second application is the parameter identification of a low order vehicle model. Most of the methods developed are based on linear, discrete-time models, being formulated as recursive, on-line techniques. A broad range of methods has been developed in the literature, see for example [29,36,50-66]. These include identification/parameter estimation techniques for purposes of vehicle control applications [50-55], fault diagnosis [56], and vehicle suspension studies [29,36,57-66]. For nonlinear systems, the model is parameterised such that it is linear-in-parameter enabling the techniques for linear system to be applied. When the model equations are linear-in-parameters, the parameter estimation problem has a standard explicit and unique solution, the well-known Least Squares regression method. Majjad [50] carried out a simulation study on the estimation of the parameters of a quarter vehicle suspension using the RLS method. Though nonlinear characteristic of the damper were involved, the parameter estimation was formulated as a linear-in-parameter case, where the parameters to be identified were the combined terms of the physical coefficients. The estimation method used was the Equation Error method where the unmeasured velocity signal was numerically obtained. In [56], the RLS method was used to estimate the parameters of a discrete-time transfer function of a vehicle suspension. A neural network was used in mapping the estimated parameters to the system physical coefficients. Zhange and Chen [59] investigated both off-line (batch processing) and on-line (recursive processing) Least Square to estimate

parameters of the dynamics of a ground vehicle dynamic. The proposed method incorporated block-pulse function for approximations of ‘unmeasured’ signals. An analytical derivation for the estimated parameter was presented with some simulation results. In [57,58] simulation studies on the parameter estimation of a ground vehicle axle subsystem and a magnetically levitated vehicle, respectively, were carried out using Instrumental Variables. Perhaps, the most severe limitation of these Least Square regression approaches is the assumption that the error signal arises as an idealised white noise process, uncorrelated with the elements of the regressor matrix. If this is not the case, the estimator may suffer from systematic bias and/or excessive variance. For example, in vehicle ride dynamics, errors arising from unmodeled mechanical vibrations have associated resonance frequencies, and the error is clearly not white noise. Also, unmodeled nonlinearities are likely to induce correlations between the error and the regressor matrix, while the errors in the regressors themselves induce bias [11,30].

The third application is the parameter identification of a complex vehicle model. This involves fitting a complex mathematical model to the real vehicle. The model can be obtained from the application of mechanical principles to idealised elements of the real system. The identification goal is then to estimate the model parameters so that then approach the real vehicle dynamics in some sense. Least squares parameter identification technique were employed in [67-73] for non-linear vehicle models as part of an advanced vehicle systems. The resulting model description is attractive as it allows both the simulation and the design optimisation of the vehicle. In [67] weighted squares of the errors, the difference between the measured and estimated vehicle responses, were used in the cost function. The optimal parameters of the

model were found using the quasi-Newton method. In [68] parameter estimation was used for fault detection of a ground vehicle in lateral motion. The paper presented resulted from a simulation study when faults of both single and combined model parameters were considered. The cost function for the optimisation procedure was formulated using both the errors in the vehicle responses and the error in the parameters. The quasi-Newton and Gauss-Newton methods were used in the parameter estimation procedure, which required knowledge of the gradient (first derivative) and the Hessian (second derivative) of the cost function. In [71-74] the parameter estimation problems was solved using several gradient-based optimisation methods such as the Gauss-Newton, the Levenberg-Marquardt and the sequential quadratic programming methods, with quadratic cost function of the response errors. The optimisation procedures shared a common feature that they all required the error to be continuous and assumed the first and second derivatives were available. However, calculation of the derivatives may be computationally expensive and/or difficult to obtain. Another disadvantage is that the optimisation algorithm often only finds local minima even if a reasonable initial guess of the model parameters was provided. An optimisation procedure that finds the global optimum parameters for the model without assumption of continuity and availability of the cost function derivatives is an alternative to overcome the parameter estimation problem.

1.5 Numerical optimisation

1.5.1 Choice of optimisation techniques

The performance of a model structure, or of parameter estimators for a given model structure, is usually rated via a cost function $J(\theta)$. Finding the best possible model then corresponds to optimizing this cost function. If the model is non-linear-in-parameters, it is difficult if not impossible to find an analytical solution. This raises the important question of what numerical optimisation algorithm should be used in searching for the best model parameters. The numerical search routines are usually iterative procedures. Starting from one or a number of initial values, a better set of parameters is generated, and this process is repeated until it is decided that the process has converged.

The gradient-based methods such as Newton-Raphson and Gauss-Newton algorithm and method of Levenberg-Marquardt are the conventional techniques. These techniques are implemented in most major libraries of scientific subroutines, and some of them are used in commercially available estimation software. In practice the convergence region of most of these methods is limited; if starting values are selected outside this region, the method will diverge. Even if there is convergence, the final result can depend upon the starting values if the cost function has a local minima. A priori information can be used to improve the starting values, but usually insufficient information is available. For example, it is difficult to give reasonable starting values for the coefficients of the transfer function of an unknown system. Another disadvantage of the gradient-based methods is that they require evaluation not only the cost function but also of its derivatives of the cost. The need to calculate the

derivatives, particularly the second derivative, can be quite time consuming. The methods also assume that the cost function is differentiable with respect to the model parameters. If this is not so, approximating method such as using finite difference method have to be used. These together can result in the gradient-based methods require a large number of evaluation of the derivatives of the cost function.

Direct search method such as the Simplex method of Nelder and Mead is useful if the derivatives cannot be easily calculated. The method does not take the local properties of the cost function into account, it can be used to optimise a noisy cost function. The algorithm however may be easily trapped in local minima and its convergence speed depends on the starting point.

Global optimisation techniques search for the global optimum points. They aim to find the best possible value of J for the cost function specified and the associated parameter vector θ . The techniques bypass the initialization problems that may arise in the gradient-based and direct search methods. A few problems among many other disadvantages of the global optimisation techniques are for example the random search, which is very simple to implement, but may fail to locate any global optimum. The second drawback is that the techniques may not be guaranteed to find the result, but at the expense of a much more complex implementation.

A relative new method named the Differential Evolution algorithm was introduced by Storn and Price [75-78] in 1996. The algorithm is simple to implement and has been successfully applied to many engineering problems [79-84]. The Differential Evolution algorithm is a type of structured random search that mimic the process of

biological evolution. The algorithm begins with a collection of parameter estimates (called a chromosome) and each is evaluated for its fitness in solving the given minimization (or maximization) task. At each generation (algorithm time-step), the most fit chromosomes are allowed to mate and bear offspring. These children (new parameter estimates) then form the basis for the next generation. The biological analogy suggests that such a procedure will be likely to lead to workable solutions for nonlinear problems. Unlike the gradient-based techniques, the Differential Evolution algorithm requires no calculation of the gradient, and is not susceptible to local minimum problems that arise with multi-modal error surfaces.

1.5.2 Motivation for the development of a new method

When solving the model parameter estimation for the problem being studied using numerical methods such as conventional (gradient-based and simplex methods) and global optimisation methods (the Differential Evolution and Genetic Algorithms) three major problems were experienced.

- The first problem was the ability and consistency (or success rate) in finding the global optimum solution. An example was in the case of the conventional gradient-based methods, which suffered the problem of obtaining local minima, and their performances were influenced by the initial starting point and the parameter bounds specified.
- The kernel of most optimisation-based parameter estimation algorithms is the simulation of the model outputs (or the prediction error), and possibly the derivatives of the objective function. Very often, the time spent in these simulations takes up most of the computational time required by the optimisation.

The second problem was thus concern with the computational time in the process of searching for the best solution. An example was in the case of the direct search method, simplex method, which has shown to have a slow convergence rate and thus time consuming.

- The third problem was concerned with the implementation of the optimisation algorithm. Numerical search algorithms usually contain several control parameters that determines the performance of the method. Tuning of the control parameters can become part of the problem, as the incorrect selection of their values would result in mis-convergence. With the global optimisation techniques such as the Genetic Algorithm, the problems experienced were the implementation difficulties and the tuning of the control algorithm parameters. The initial investigation into the Differential Evolution found that the method was easy to implement and was able to find the global optimum solution. However, it has a slow convergence rate due to its ‘discontinuous’ behaviour, and the problem of expensive computational time due to the simulation of the model, which has not been dealt with.

The problems have led to the development of a new numerical optimisation method that has become the central part of this thesis work.

1.6 Thesis aims and Outline

According to the problems experienced when solving the model parameter estimation for the problem being studied using the conventional and global optimisation methods, as described in the last section, the specific aims for this thesis were set as follows;

- 1. To develop a practical optimization technique, which is to be used in the parameter estimation of nonlinear systems whose models are nonlinear-in-parameter, or when obtaining the objective function is computationally expensive. To overcome the problems experienced, the algorithm should fulfill the following requirements;
 - Have the ability to deal with premature termination due to local minima and mis-convergence problems due to non-differentiable, nonlinear and multi-modal objective function.
 - Have good convergence properties; this means, first, consistent convergence to the global optimum in consecutive independent runs, and secondly, ability to cope with computation intensive objective function.
 - Ease to implement and use, with a few control parameters that are easy to choose.
- 2. To validate the newly developed algorithm and examine its performance by comparing the results with other techniques such as the conventional (gradient-based and direct search methods) and the global optimisation methods (the Differential Evolution algorithm).
- 3. To apply the concept of system identification and parameter estimation to a number of practical problems relating to vehicle ride dynamics using the proposed algorithm.

The rest of the thesis is organized as follows.

Chapter 2 describes the proposed discrete variable Hybrid Differential (dvHDE) algorithm in details. First, the motives behind the development of the method and why it is based on the Differential Evolution (DE) algorithm are given. The dvHDE method shares a common algorithm structure with the original DE method, however, there are fundamental differences between them. The common and different functional features between the two methods will be explained one by one, and guidance for setting the algorithm control parameters is also provided. Three numerical examples are considered to demonstrate and validate the performance of the dvHDE method. The examples are chosen from the parameter estimation problems involving different types of parameters. The first example is a mass-spring-damper problem when all of the parameters involved are continuous. The second example is a gear train design problem where the objective of the optimisation task is to choose the number of gear teeth in the gear train configuration such that to obtain the maximum output torque for a given input torque. The parameters involved are thus all integers. The third example considers the optimisation of a coil spring, where the parameters involved are a mixture of integer, discrete and continuous variables. The investigation on the chosen examples enables possible problems during the search, such as premature termination due to local minima and mis-convergence, to be illustrated and explained explicitly. The performance of the dvHDE method is compared with a number of other techniques as reported in the literature. The suggestions and conclusions regarding the performance and potential use of the dvHDE method are made at the end of the chapter.

The numerical examples selected for illustrative purpose in Chapter 2 are nonlinear programming problems, however, they do not involve experimentation. Chapter 3, 4 and 5 aim to investigate further the performance of the dvHDE method when it is employed in the system identification and parameter estimation process of more complicated problems from experimental data. This is carried out by investigating three practical case studies relating to the dynamics of vehicle component, subassembly and a full vehicle. In the analysis of results for each case study, the performance of the dvHDE method is compared to the conventional methods, the gradient-based (GB) and Downhill Simplex (DS) methods, and the original Differential Evolution (DE) method.

The first case study considered in Chapter 3 is the parameter estimation of an automotive damper, one of the most important components in modern vehicle suspension systems. The task involves modelling and experimentation work on an automotive damper for a wheeled passenger car when the frequency range of interest is 0.5-30 Hz. The objective of the exercise is to select the best model among a set of models considered for the test damper. The chapter provides a brief discussion on the modelling of a damper, and then, proposes a set of damper models derived from combinations of spring and viscous damping elements. The estimation is formulated as a single objective optimisation problem where the objective function is formulated from the difference between the measured and modelled damper forces. The estimation results are discussed and conclusions regarding the performance of the dvHDE method and modelling of the automotive dampers are made. In addition, an investigation into the parameter estimation of an adjustable damper is presented in Appendix A. The additional investigation is aimed to further prove the use and

performance of the dvHDE method when the damper model is more complicated (when it is modelled from physical details, not from a combination of idealized damping and spring elements, such that the model is capable of describing damper behaviour better).

A full wheeled vehicle represents a complex system with many degrees of freedom. In the development to the estimation of its parameters, a simplified system, ‘single wheel station’, provides an ideal system to test the identification and parameter estimation method. The ‘single wheel station’ is a simplified system that represents the vertical dynamics of a quarter of a vehicle, and is useful in the vehicle suspension investigations. In Chapter 4, system identification and parameter estimation process is carried out in both the time and frequency domains to obtain a model and its parameters for the single wheel station. The single wheel station, whose parameters represent an independent suspension of a medium-sized family car, is modelled based on a quarter vehicle model. The chapter discusses how a quarter vehicle model can be derived from a full vehicle model, and investigates how the variation of each model parameter influences the model frequency responses using Bode plots, a technique mainly used for linear time variant systems. A method of obtaining the model frequency responses is developed when nonlinear elements are incorporated into the model, the method is later used in the parameter estimation in the frequency domain. The estimation is formulated as a multi-objective minimisation of the difference between the system measured and modelled responses; for the estimation in the time domain, the responses are five time-history signals, and in the frequency domain, the responses are the gains and phases of the system outputs relative to the given input.

The qualities of fit of different models are analysed and the performances of different numerical search methods are compared, before conclusions are drawn.

The parameter estimation technique developed in Chapter 4 is extended to a wheeled vehicle problem in Chapter 5. The vehicle under the investigation is a Landrover 110. The dynamic behaviour of the vehicle is complicated due to the coupling of the motions in different modes, and nonlinearities due to suspension linkages and compliance. As a consequence, the initial investigation in this chapter is focus on the vehicle dynamics in bounce and pitch motions. The vehicle model parameters are estimated by fitting the modelled frequency responses to the system frequency responses. The estimation is formulated as a multi-objectively optimisation problem, where the objective function is a weighted least squared of the differences between the gains and phases obtained experimentally and from the model using the technique developed in Chapter 4. Due to limited time, only the parameter estimation in the frequency domain is carried out, with a vehicle bouncing and pitching model. For the parameter estimation in the time domain and more complicated vehicle model that usually requires complicated commercial computer software to generate the vehicle equation of motion are left for future work. The main interest here is again the parameter estimation task, and the use of the dvHDE method.

Finally, a summary, the conclusions and some proposals about future research based on aspects discussed in this thesis are given in Chapter 6.

CHAPTER 2

THE DISCRETE-VARIABLE HYBRID DIFFERENTIAL EVOLUTION

This chapter aims to give a detailed description of the proposed discrete variable Hybrid Differential Evolution (dvHDE) method. The dvHDE method is based on the Differential Evolution (DE) method of Storn and Price [75-77]. In section 2.1, the original DE method is introduced and its performance at solving some engineering problems as reported in the literature are discussed, followed by the motives behind the development of the dvHDE. In sections 2.2 to 2.6, the main functional features of the dvHDE method are explained.

The performance of the dvHDE method is controlled by several parameters, suitable values for these parameters are given in section 2.7. To demonstrate and investigate the performance of the dvHDE method against the original DE method and some other techniques, the solutions to three numerical optimisation problems are considered. The first example is a three-dimensional optimisation problem relating to dynamics, in which the parameter involved are all continuous variables. The other two problems involve the optimisation of integer, and a mixture of integer, discrete and continuous variables in engineering design applications. The results for this are presented in section 2.8, and problems during the search for the optimum solutions such as premature termination and mis-convergence are illustrated explicitly. Finally, from an analysis of the estimation results, suggestions and conclusions regarding the performance of the dvHDE method are given in section 2.9.

2.1 Introduction

2.1.1 Introduction to Differential Evolution

This section begins by giving a brief description of the DE method on which the dvHDE is based, then, discusses the performance of the DE method at solving some engineering problems as reported in the literature. For further information on the DE method, the reader is referred to references [75-78].

Differential Evolution (DE) is a population based and stochastic function minimizer (or maximizer), whose simple yet powerful and straightforward to apply features make it a very attractive numerical optimisation method. The DE was developed by Storn and Price and was originally designed for function optimisation of continuous space problems. The general structure of the algorithm consists of the followings; representation and initialisation, evaluation, recombination and natural selection based on the principle of survival of the fittest.

Considering a three-dimensional parameter estimation problem, where a possible solution contains three parameters to be identified. The representation issue, also called the encoding procedure, is concern about how the algorithm construct the solution and its components. The binary, integer and real number encoding techniques are typical examples; each component (each of the three parameter) of a solution is represented as a string of 0 and 1, an integer and a real number in the binary, integer and real number encoding techniques, respectively. The initialisation is concern with the generation of the initial solutions. Each solution is assigned a value in the evaluation phase enabling it to be rated as to how good it is for the problem (fitness of

that particular solution). A set of the solutions is called a population. The production of an improved population is the kernel of the optimisation method, this is carried out in the recombination phase. The outcome from the recombination procedure is a new population, in addition to the current population. The members in the two sets of populations compete with one another in the selection phase to form the next set of solutions. The selection is based on the principle of survival; this means the solution, which is higher rated (fitter), is likely to be included in the next set of solutions, as the estimation process continues.

The important and fundamental idea behind the DE method is a scheme for which it generates the trial parameter vectors. The DE uses vector differences for perturbing the vector population. The general concept begins with a pair of parameter vectors randomly selected from the current population, their difference is multiplied by a constant, the result is then added to a third parameter vector to form a trial parameter vector. The operation is carried out in parallel for all members in the population so that a population of the trial parameter vectors of the same size as the current population is generated. An example of this process when generating a trial parameter vector, \mathbf{v}_4 , of a simple two-dimensional minimisation problem is given in Fig. 2.1.1. In this particular example, \mathbf{v}_3 is perturbed by $0.5(\mathbf{v}_2 - \mathbf{v}_1)$ to obtain \mathbf{v}_4 . Here, we refer to the current population of parameter vectors as parents, and the trial parameter vectors as offspring. The offspring \mathbf{v}_4 is therefore better than its parent \mathbf{v}_3 as justified by $J(\mathbf{v}_4) < J(\mathbf{v}_3)$, where J is a measure of the degree of satisfaction to the minimisation criterion.

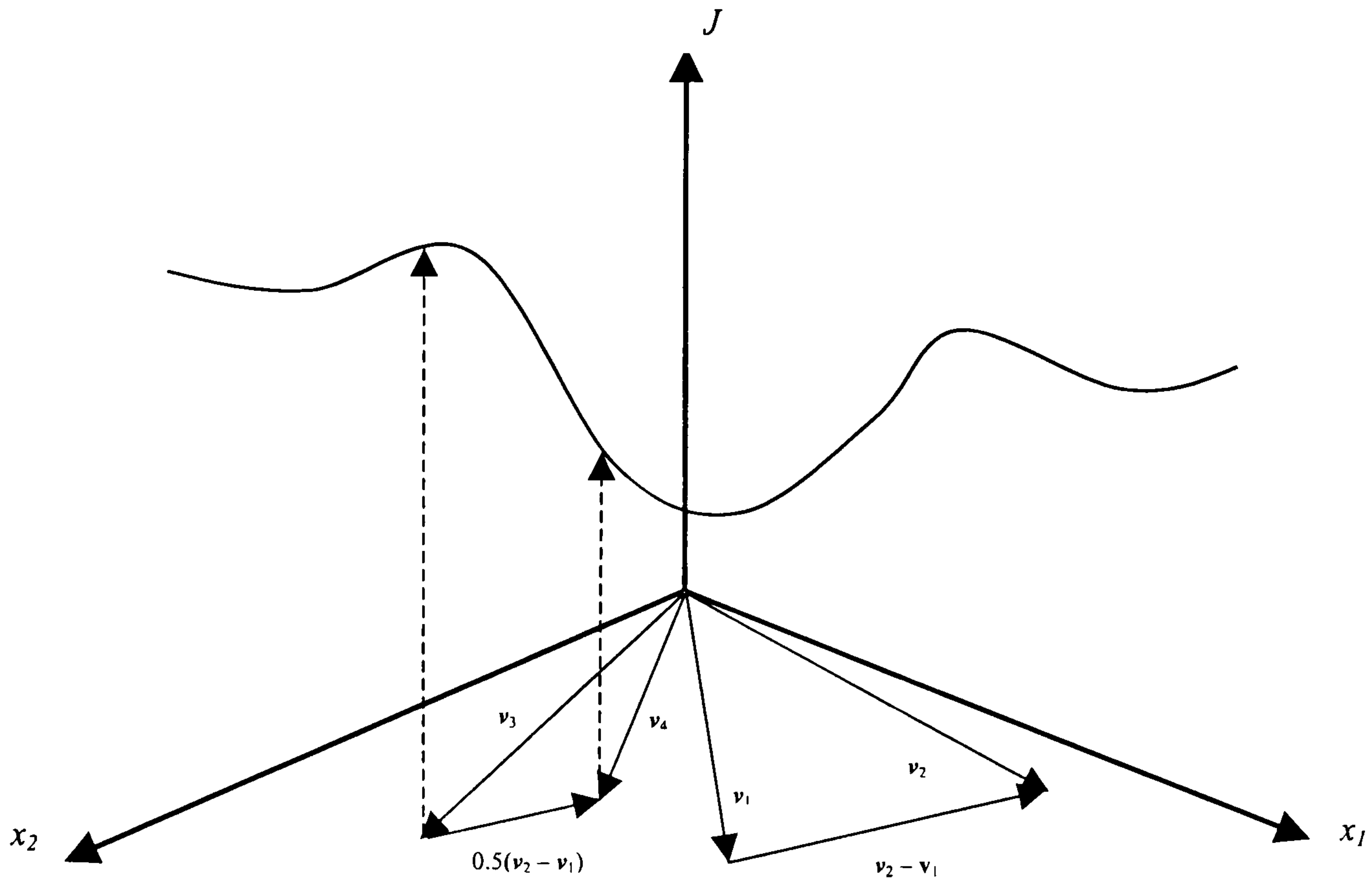


Fig. 2.1.1 An example of the process of generating a trial parameter vector, v_4 .

The selection strategy of the DE method is a one-to-one competition between the parent and its corresponding offspring. Between the two, the one which is higher quality in terms of fitness to the specified criterion is chosen to be a member of the next generation. This means, the offspring will replace its corresponding parent if its fitness is higher, otherwise the parent is retained. From Fig. 2.1.1, v_4 is ‘fitter’ and is selected for the next generation. Starting with the initial population, the algorithm evolves towards successively better regions of the search space by means of recombination and selection. The best individual obtained when the algorithm is terminated is then taken as the optimum solution to the problem at hand.

In many real world applications, engineers and researchers commonly employ conventional gradient-based (GB) search methods when solving optimisation

problems. Typical examples are the sequential quadratic programming, quasi-Newton, Gauss-Newton and Levenberg-Marquardt methods. Alternatively a Genetic Algorithm-based (GA) approach may be used. The GA does not require information about the derivatives of the objective function, and provides multi-point search, robustness and global convergence in a complex search space. The DE method, however, has been shown to be a further improvement over the GA-based method in that the DE is able to find a better solution and is capable of doing so more consistently, see for example in [79]. The DE method won 3rd place at the First International Contest on Evolutionary Computation (ICEC) held in May 1996, and it turned out to be the best genetic type of algorithm for solving the real-valued test function (the first two places were given to non-GA type algorithms that are not universally applicable but solved the test-problems faster than the DE [78]). When comparing the DE with conventional GAs technique, they share a common general structure; initialisation, evaluation, recombination and natural selection based on the principle of survival of the fittest, but have the following basic differences. Firstly, the DE uses real-number encoding to represent an individual in the population, instead of using strings of zeros and ones like conventional GAs. The reproduction, which is the process of generating a new population of the solutions from the current population, in the DE method is performed on real-floating-point numbers. Secondly, there is a difference in the way the DE generates a perturbation to the current population and in the selection strategy when producing a population for the next generation. These two points will become clear when they are explained in sections 2.3 and 2.5.

Since introduced, the DE method has been successfully applied to many engineering applications [79-84]. The 'DE community' has been growing since its early years of

1994-1996, and more researchers are working on and with the DE method. In [80] the DE method was applied to identify the model parameters of an automotive engine mount. The component was modelled using the Freudenberg hydromount model that contained a set of highly non-linear piecewise-continuous differential equations. In [81] the parameters of the Bouc and Wen model, a widely used model in the field of hysteretic or memory-dependent vibrations, were estimated using the DE method. The model structure included internal states and nonlinear terms. The DE method has been successfully applied in all these difficult optimisation problems. In general, optimisation of nonlinear systems are nonlinear programming problems. When formulating the objective function in nonlinear programming, the model parameters usually are assumed to be continuous. However, it is common in practice, such as in engineering design work, that the model parameters can only have integer or discrete values. For example, the design of a vehicle suspension coil springs where the outer diameter of the spring can be any value, thus it is a continuous variable, while the available steel wire diameters is discrete, and number of spring coils is an integer. The DE method is capable of handling mixed integer, discrete and continuous optimisation problems, for example the study by Lampinen and Zelinka in [82].

2.1.2 Motivations to the development of the dvHDE method

This research study has modified the original DE method and developed an approach where all integer, discrete and continuous model parameters are treated in the same unified way. All of the parameters involved in the optimisation are treated as discrete variables. The proposed method divides the entire search space into intervals whose

increments satisfied the minimum accuracy required. It also includes two additional operations that improve convergence speed and avoid possible premature termination problems. The two additional mechanisms, namely the migration and acceleration operations, will be explained in section 2.6. Since the method developed in this work treats all variables as discrete variables and employs additional features in addition to the normal mechanisms of the original DE method, it has been named ‘the discrete-variable Hybrid Differential Evolution (dvHDE)’ method. The motives behind the development of this dvHDE method and why it is based on the DE method are summarised below.

First, previous investigation of the gradient-based (GB) method in [85,86] on optimisation problems relating to vehicle suspension applications showed a slow convergence rate and was very time-consuming. Initial investigation of the DE method on the same problems showed that, for several consecutive iterations, the DE method was unable to or only made a slow progress from one region of the search space to a better one. The DE convergence speed appeared to show a ‘discontinuous’ characteristic. This problem will be illustrated later in section 2.8. A numerical search method that improves the convergence speed over the GB method, and reduces or eliminates the discontinuous characteristic of the DE method is certainly desirable. In addition, the DE method has a fast convergence rate. However, fast convergence rate also leads to a greater possibility of getting trapped in local minimum points or mis-convergence. The original DE method employs no mechanisms to deal with these problems. The search procedure could be trapped in one of several possible local minimum points and fail to find a global optimum. Additional mechanism needs to be introduced if those problems are to be overcome.

Secondly, the DE method is a simple and efficient algorithm for solving problems with parameters that are not only continuous variables but may also be integer or discrete. The dvHDE method, which is based on the DE method, employs the same main working mechanism. This should hopefully give a guarantee that the advantages of the DE method are also possessed by the dvHDE method. The actual performance of the dvHDE method must however be investigated to verify its behaviour. This is best done by applying the algorithm to different real-world applications. In this work, the investigation has been limited to focusing on the application of the dvHDE method to parameter estimation for vehicle dynamics problems. However, the algorithm can be applied to other real-world applications, particularly to the optimisation of nonlinear systems that involves computationally time consuming when calculating the objective function. Such cases will be further discussed later in this sub-section. The dvHDE method has so far been applied to optimisation of small and medium sized problems where the number of parameter to be identified is less than thirteen, problems with a larger number of parameters could be considered in future work.

Thirdly, considering the typical optimisation problems where the parameters involved in the objective function are continuous variables. It is usually true that each variable has a given level of precision, it is therefore acceptable to ‘re-treat’ continuous variables to discrete variables with an increment of a sufficiently small size so that the levels of precision are maintained. The search space by this method is converted from a continuous to a discrete one. Since integers are discrete, optimisation problems involving integer variables can directly be treated as discrete variable problems.

Problems of optimisation, which possibly involve integer, discrete and continuous variables, or a mixture of the three type of variables, can now be treated in a unified manner.

Lastly, the time required to complete one cycle of the parameter estimation process of a model selected for the system of interest, is dependent upon two main factors. One factor is the performance of the numerical search method used to optimise a specified objective function. The other factor is the number of variables in the objective function. The objective function for parameter estimation problems usually requires experimentally obtained system responses and system simulated responses. It may also require other variables such as weighting matrices used in the Weighted Least Square and Maximum Likelihood estimators of multiple output systems. There are many cases, see for example [1, 80-84], where the objective function and its derivatives may be computationally expensive to obtain. The response of a complex model, which contains many parameters and involves large amount of numerical calculation such as differentiation and integration, often requires a long time to compute. The responses may be either time-domain or frequency-domain responses. In the case of a parameter estimation in the frequency domain, particularly for non-linear systems, the most time consuming part may result from the technique employed to obtain the model frequency responses including the amplitude ratios and relative phases of the system outputs relative to that of the system input. If not known prior to the estimation process, calculation or determination of other variables such as the weighting matrices in the objective function can also be an important contribution to the total time consumed. Therefore, when dealing with parameter estimation, particularly with nonlinear systems that often require high model complexity in order

to closely mimic the system behaviour, the time scale to complete the estimation task can take significant amount of time.

When the objective function is computationally expensive to obtain, either due to model complexity or the determination of variables, the requirement is to perform the objective function evaluation a minimum number of times. The number of objective function evaluations may be reduced by, first, using a smaller population size, or setting-up the algorithm control parameters for a faster convergence speed, or introducing an additional mechanism that helps speed up the convergence rate. These can also result in the search being trapped in local minima or mis-convergence.

The second is to avoid re-evaluation of repeated solutions, bearing in mind that the number of repeated solutions depends on the problem size and tends to increase as the search progresses towards the termination. If all parameters involved are all treated as discrete variable, and if every possible solution has an identifier, a reduced number of function evaluations is achieved by not re-evaluating, but checking the associated objective function values from previous occurrence. The benefit of doing this depends on the size and complexity of the optimisation problem; for the problems investigated in this work the benefit was significant. For problems with larger size, the number of repeated solutions may be small relative to the total number of possible solutions in the search space, however the benefit may become significant when it is computationally expensive to obtain the objective function.

The benefit of checking for repetition of the solutions and avoiding re-evaluation may increase when information (the identifiers and their associated fitness values) during

the tuning of the algorithm control parameters is passed on from run to run. All numerical search methods have a set of control parameters that dictate the efficiency of the methods. In population-based methods such as GAs and DE, the control parameters include pre-assigned constants of population size, recombination rates and maximum number of generations (or iterations), and their optimum values are usually tuned by trial-and-error. The number of repeated solutions increase, and with the dvHDE method, the number of function evaluations thus reduces if the information during the turning is passed on from run to run. Once one is satisfied with a set of the control parameters after the trial-and-error process, one usually runs the optimisation algorithm several times before arriving at the final solution that statistically reliable. With the dvHDE method, the number of repeated solutions may be increased, thus the number of function evaluations further reduces, if the information (the identifiers and associated fitness values) is passed on from previous run to the present and future run. All these are not possible for the DE or GA methods as no such information (the identifier and associated fitness value) are available. The total benefit of checking for repetition of solutions has two limitation that must be born in mind; problems size, and the limitation due to the amount of computer memory used in storing those identifier and their associated fitness values when using a single CPU. The later may be dealt with by better organising the memory usage or using network PC. Due to limited time, the work in this thesis does not include an investigation into this possibility, and is left for future work.

The ultimate aim of developing the dvHDE method in this work is to reduce the number of function evaluation and thus improve convergence speed without sacrificing the consistency of finding the global optimum solution. The algorithm is

specifically aimed at non-linear systems, where optimisation of the objective function can be computationally time-consuming. The basic principles of dvHDE are taken from the original DE method, the fundamental difference being the encoding technique used; the dvHDE method uses integer encoding rather than real numbers. Furthermore, the dvHDE method has two additional operations, namely the Acceleration and Migration. These two operations are performed to attempt to increase convergence speed and prevent premature convergence by maintaining high population diversity. In the following sections, the basic operations of the dvHDE method are described.

2.2 Representation and initialisation

This section describes how a solution for a parameter estimation problem is encoded and how a population of the initial solutions is generated by the dvHDE method.

In many real world applications, one requires to know the parameter values of the system of interest only to a certain level of accuracy. For example, it may be sufficient to know the mass of a system of interest accurate or precise up to ± 0.01 kg, or the stiffness to ± 1 N/m, or the damping coefficient to ± 0.1 Ns/m. From the a-priori knowledge of the system, one usually can sensibly decide upon an initial range for each parameter. With the precision levels set and the parameter lower and upper bounds defined, the number of parameter values is finite and the parameter space limited. Now the parameter values exist only at discrete points in the entire parameter space. However this does not mean that the dvHDE leaves out some parameter

resulting in a failure to find the optimum solution, since parameter precision can always be increased.

Considering again a three-dimensional parameter estimation problem where all of the parameters involved are continuous variables. The dvHDE method make the search space finite by ‘discretising’ each of the three continuous variables according to their minimum required precision. The discrete version of the j^{th} -parameter can be obtained using equation (2.1);

$$\theta_{i_{\theta,j}} = \theta_{\min,j} + (i_{\theta,j} - 1) * \theta_{tol,j} \quad (2.1)$$

$$j = 1, 2, 3, \dots, n_j \quad \text{and} \quad i_{\theta,j} = 1, 2, 3, \dots, n_{\theta,j}$$

Where n_j is number of the parameter to be identified, $i_{\theta,j}$ is the associated index of each discrete value point in ascending order, and $n_{\theta,j}$ is the number of discrete value points of the j^{th} -parameter, which can be easily computed from the user specified parameter precision, $\theta_{tol,j}$, and the parameter lower and upper bounds, $\theta_{\min,j}$ and $\theta_{\max,j}$, respectively (Noting that $\theta_{i_{\theta,j}} = \theta_{\min,j}$, when $i_{\theta,j} = 1$, and $\theta_{i_{\theta,j}} = \theta_{\max,j}$, when $i_{\theta,j} = n_{\theta,j}$ and $n_j = 3$ for the three dimensional problem). The dvHDE converts the original continuous parameter space to a discrete one using equation (2.1), and it is now the index of the discrete value points that represent the solution space of the problem at hand.

Two ways of generating/representing the search space have been considered in this work, a brief description of each method is given here. The first method is to generate

the search space with a regular parameter grid, and store all possible solutions. A search is then performed on this discrete space. However, this method is only suitable for small problems where the number of parameters to be identified is small and values of $\theta_{tol,j}$, $j = 1, 2, 3, \dots, n_j$ are large. When the number of parameters is large and/or the parameter precision required is high, i.e. values of $\theta_{tol,j}$, $j = 1, 2, 3, \dots, n_j$ are small, the search space generated become vary large. In such a case the search space must be partitioned if the problem is to be solved efficiently [87-88]. In the second method, the existence of the discrete solution points in the search space is virtual. The search space is not stored in computer memory. A solution or a point in the search space is generated only when it is required. After evaluation, its identifier and fitness values are stored and used for checking/avoiding function re-evaluation, current parent and offspring populations are overwritten every generation. It is this second method that has been employed in this work.

Having discretised the parameter involved using equation (2.1), a *gene*, as it is commonly referred to in Evolutionary Algorithm (EAs) and GAs language, is then an index of the parameter at a discrete point, and a *chromosome*, which is an alternative word for a possible solution to the problem, contains genes which are all integer. For the three-dimensional optimisation problem, a chromosome thus comprises of three genes. The dvHDE further benefits from using discrete versions of the variables and from using integer-encoding technique by augmenting one extra gene into the chromosome. This extra gene is an integer number identifying the chromosome in the entire parameter space. Its purpose is to determine whether a chromosome has been evaluated for fitness, thereby avoiding re-evaluation of the repeated chromosomes

(whose number depends on the size of the problem). This extra gene is only used for avoidance of re-evaluation, it is not involved in any recombination operations.

When implementing the dvHDE algorithm, a gene is represented by x_{ij} as an element in a matrix ${}_gX_{i,j}$, where ${}_gX_{i,j}$, $i = 1, 2, \dots, n_i$ and $j = 1, 2, \dots, (n_j + 1)$ denotes the population at generation g^{th} , n_i is the population size, and $(n_j + 1)$ denotes the number of parameters to be identified plus the identifier gene. Thus at generation g^{th} , there are n_i individuals or chromosomes, and a chromosome is comprised of $(n_j + 1)$ genes. Considering the current population, the i^{th} -row of ${}_gX_{i,j}$ represents the i^{th} individual, the j^{th} -column of ${}_gX_{i,j}$ represents the j^{th} gene of the corresponding individual.

It is the general practice that the initial population are randomly selected between the lower and upper parameter bounds so that they uniformly cover the entire range specified. The dvHDE method generates its initial population by generating n_i random numbers between 1 and $n_{\theta,j}$, $j = 1, 2, \dots, n_j$ (where n_i and $n_{\theta,j}$ are population size and the number of discrete value points of the j^{th} -gene, respectively). An identifier for each chromosome of the generated initial population is then calculated and assigned in the $(j+1)^{th}$ column of ${}_{g=0}X_{i,j}$.

2.3 Mutation

Like the original DE method, the dvHDE method has two important operations for the production of offspring. They are the mutation and crossover operations. This section

describes the mutation operation, and the crossover operation will be explained in the next section. The mutation operation is the central working part of both the DE and dvHDE providing a distinct mechanism by which the next search points in the search space are generated. This operation is performed at the chromosome level, it begins with two chromosomes ${}_gX_{i_1,j}$ and ${}_gX_{i_2,j}$, which are randomly selected from the current population. The mutated chromosomes ${}_g^mX_{i,j}$ are then obtained according to equation (2.2) given below.

$${}_g^mX_{i,j} = {}_gX_{i_3,j} + \Theta[\mu_m ({}_gX_{i_1,j} - {}_gX_{i_2,j})] \quad (2.2)$$

$$i_1, i_2 \in \{1, 2, 3, \dots, n_i\} \quad \text{and} \quad i_1 \neq i_2$$

Where ${}_gX_{i_3,j}$ is a third chromosome (this may be the best chromosome found from previous generations, or just another randomly selected chromosome from the current population, i.e. $i_3 \in \{1, 2, 3, \dots, n_i\}$), $\Theta[\cdot]$ is a rounding operator that rounds the elements to their next nearest integers, and μ_m is the mutation control factor that acts as a scale to which the vector difference $({}_gX_{i_1,j} - {}_gX_{i_2,j})$ is added to ${}_gX_{i_3,j}$. In the original DE, μ_m is fixed and is suggested to be in the interval of (0,1). This value can be set by the user to ensure the fastest possible convergence for a particular problem. The value of μ_m for the dvHDE method is also in the interval of (0,1). In addition, three options for setting the values of μ_m have been implemented in the algorithm. The first option is a specified value between (0,1) by the user, which is held constant for all generations through out the search. The second option is to randomly generate a number between (0,1). This value is re-generated, so that it is different from one

generation to another. The third option is a matrix of size $n_i \times n_j$, the elements of this matrix are random numbers between (0,1), which are re-generated for each generation. The second and third options provide alternative set-ups to the first option that may result in a better exploration of the search space. The applications of the dvHDE method to the problems reported in this work all have employed the first options, that is μ_m is a fixed number between (0,1). The second and third options showed some good results when experimented with however they have not been fully investigated.

Since the dvHDE method employs an integer-encoding technique, the mutation operation may produce illegal chromosomes. This is when at least one of the genes in a chromosome does not belong to the search space. An example is when a gene, which is the index, is zero or negative. The dvHDE repairs this by replacing an illegal gene with a randomly selected feasible one.

2.4 Crossover

The crossover operation is performed at the gene level, its objective being to increase local diversity of the population. The mutated chromosomes ${}^m_g X_{i,j}$ obtained from equation (2.2) and the parent chromosomes ${}_g X_{i,j}$ are used to generate the offsprings, ${}^c_g X_{i,j}$. Each gene of the chromosomes in ${}^c_g X_{i,j}$ is chosen from the corresponding genes in ${}^m_g X_{i,j}$ and ${}_g X_{i,j}$ using the binomial distribution. For example, the j^{th} -gene of the i^{th} -chromosome of the offspring population is chosen from the j^{th} -genes of the i^{th} -

chromosomes of the parent and the mutated chromosomes, ${}_gX_{i,j}$ and ${}_mX_{i,j}$ respectively, and between the two competing genes, the possibilities of getting selected and not getting selected is 50/50 percent. The crossover operation is therefore expressed as,

$${}_g^cX_{i,j} = \begin{cases} {}_gX_{i,j} & \text{if a random number is } > \mu_c \\ {}_mX_{i,j} & \text{otherwise} \end{cases} \quad (2.3)$$

$$j = 1, 2, 3 \dots n_j \text{ and } i = 1, 2, 3 \dots n_i$$

where μ_c is the crossover factor and is an additional control parameter used in the dvHDE. The crossover factor $\mu_c \in [0,1]$ is fixed and set by the user, guidance to the selection of its value will be given in section 2.7.

2.5 Evaluation and Selection

Prior to the selection, each offspring obtained from the crossover operation has to be assigned a fitness value, this is done at the evaluation stage. The dvHDE first checks whether an offspring needs to be evaluated. If required, evaluation is executed otherwise the fitness value is referred from the previous occurrence in the past generations of that individual.

The population size remains constant from generation to generation. Like the DE, the dvHDE method employs a greedy strategy in its selection operation. This means that

the individual with better quality in the sense of fitness to the specified criterion is favoured for the next generation. The one with a worse fitness is excluded from the population. The selection operation of the dvHDE is a one-to-one competition between each pair of parent and offspring. This competition means that the parent is replaced by its offspring if the fitness of the offspring is better than that of its parent. On the other hand, the parent is retained in the next generation if the fitness of the offspring is worse than that of its parent.

The information about the best individual of the current generation and the best individual so far found from the previous generations are collected and updated during this stage. Then, the selection operation may be expressed mathematically as

$${}_{g+1}X_{i,j} = \operatorname{argmin}\{ J({}_gX_{i,j}), J({}_g^cX_{i,j}) \} \quad (2.4)$$

$${}_g^b x_j = \operatorname{argmin}\{ J({}_{g+1}X_{i,j}) \} \quad (2.5)$$

$$j = 1, 2, 3 \dots n_j \text{ and } i = 1, 2, 3 \dots n_i$$

where *argmin* means the argument of the minima; in equation (2.4), between each pair of the two competing chromosomes in ${}_gX_{i,j}$ and ${}_g^cX_{i,j}$, *argmin* returns the chromosomes which is lower value of *J*, where *J* is a function that assigns fitness measure to each chromosome, noting that the fitter the chromosome the lower the value of *J*. In equation (2.5), *argmin* returns the best chromosome, ${}_g^b x_j$, whose value of *J* is the smallest.

2.6 Acceleration and Migration

The problems experienced during the application of the DE and gradient-based (GB) methods were slow convergence speed and premature termination. The acceleration and migration operations have been included in the dvHDE to overcome those problems. The acceleration operation is used to speed up the convergence. However, fast convergence might lead to a higher probability of obtaining a local optimum. In the DE method where the mutation factor μ_m is fixed in the interval of (0,1), a higher probability of getting a premature termination results from a fast decreasing of the difference vector as observed from equations (2.2) and (2.3). As generations progress, eventually the population will closely cluster at one point in the search space, usually around the best chromosome, ${}^b_g x_j$. When encountering a premature termination problem, the migration can be used to escape from this local point. A new population of offspring chromosomes is randomly migrated away from the best chromosome, ${}^b_g x_j$, to the whole search space. Accordingly, the diversity of the offspring population, which competes against the parent population in the selection operation, can be retained by such a migrating operation.

Acceleration

The problem experienced when applying the DE method to solve optimisation problems was that the best fitness did not descend continuously from generation to generation. There were also several consecutive generations that the change in best fitness was small or none. An acceleration operation has been incorporated in the dvHDE algorithm in an attempt to overcome this ‘discontinuous’ type of descent and

speed up the convergence rate thereby increasing the rate of reduction in the best fitness value. The acceleration operation is basically a local search routine such as the steepest descent or simplex methods. In this work, the Downhill Simplex search (DS) of Nelder and Mead [89] has been employed in the acceleration operation. The DS is a fast local search method that operates on a (n_f+1) -sided polygon, called the simplex. It is a direct search, no gradient information of the objective function is required. Its working principles rely on the reflection, expansion, and contraction of the simplex. The acceleration operation should achieve its aims without making the dvHDE prematurely converge to a local minimum. The DS quickly finds better individuals in the local downhill directions, while the recombination and selection operations use the diversity of the population to decide whether to include the improved individuals for the next generation. In this work, the improved individual from DS acceleration replaces the worst member of the current population. To maintain the population diversity, the DS is performed for only a limited small number of iterations, and the acceleration operation is only carried out when the best fitness of the current generation is less than that of the previous generation by a specified percentage.

Migration

The speed of convergence can be improved by the acceleration operation, however, faster descending usually results in the algorithm being trapped in local minima or mis-converge, and performing the acceleration operation too often can cause the population to cluster around one point of the search space.

The migration operation is included in the dvHDE as a mean of escaping a local minimum and preventing premature convergence. In this operation, a new population is generated and used for the next generation. The new j^{th} -gene of the i^{th} -chromosome may be produced based on the best chromosome, ${}^b_g x_j$, as follow;

$${}_{g+1}X_{i,j} = \begin{cases} {}^b_g x_j + \Theta[\delta_1 (1 - {}^b_g x_j)] & , \text{if } \frac{{}^b_g x_j - 1}{n_{\theta,j} - 1} < \delta_2 \\ {}^b_g x_j + \Theta[\delta_1 (n_{\theta,j} - {}^b_g x_j)] & , \text{otherwise} \end{cases} \quad (2.6)$$

$$i = 1, 2, 3, \dots, n_i \quad \text{and} \quad j = 1, 2, 3, \dots, n_j$$

where $n_{\theta,j}$ is as in equation (2.1) belonging to the j^{th} -gene, δ_1 and δ_2 are uniformly distributed random numbers in the range of (0,1) and $\Theta[\cdot]$ is the rounding operator that rounds the elements to the next nearest integers.

The migration operation is executed only when the population diversity measure is lower than a specified value. The population diversity is a statistical measure that indicates how population scatters in the entire search space. It may be computed as the average of the number of genes that are different from the corresponding genes of the best individual by, say, 5 %. Mathematically it may be expressed as

$$\rho_m = \frac{\sum_{i=1}^{n_i} \sum_{j=1}^{n_j} \eta_{i,j}}{n_j (n_i - 1)} < \varepsilon_1 \quad (2.7)$$

$$\eta_{i,j} = \begin{cases} 1 & , \text{if } \left| \frac{{}_g X_{i,j} - {}^b x_j}{{}_g x_j} \right| > \varepsilon_2 \\ 0 & , \text{otherwise} \end{cases} \quad (2.8)$$

$$i = 1, 2, 3, \dots, n_i \quad \text{and} \quad j = 1, 2, 3, \dots, n_j$$

ε_1 and ε_2 are the desired tolerance for the population diversity and the gene diversity with respect to the best chromosome, ${}^b x_j$.

A flowchart of the dvHDE's operations is provided in Fig.2.6.1. The algorithm starts with a randomly generated population of size n_i -row x n_j -column, which is uniformly selected between the lower and upper bounds. The mutation operation is carried out next to produce a perturbation to the initial population that originates the algorithm evolution from one region of the search space to another. The result from the mutation operation is a population of mutated chromosomes, some of which may contain illegal genes. The algorithm therefore checks for illegal genes. If there is, they are repaired. The crossover operation then follows and the outcomes are a population of the offspring. The identifiers are computed and assigned to the $(j+1)^{th}$ genes of the offsprings' chromosomes. The algorithm then checks for repeated occurrence of the offspring. If there is a repetition, the chromosome fitness value is used from its last occurrence. Otherwise, the evaluation of the fitness function will be carried out. The algorithm then performs the selection operation selecting between the parents and their corresponding offspring based on the fitness values. The next two operations are the acceleration and migration, which are carried out only when their initiation criteria are met. The algorithm has now completed one generation of the evolution process. It will evolve to the next generation until the specified termination criterion has been reached.

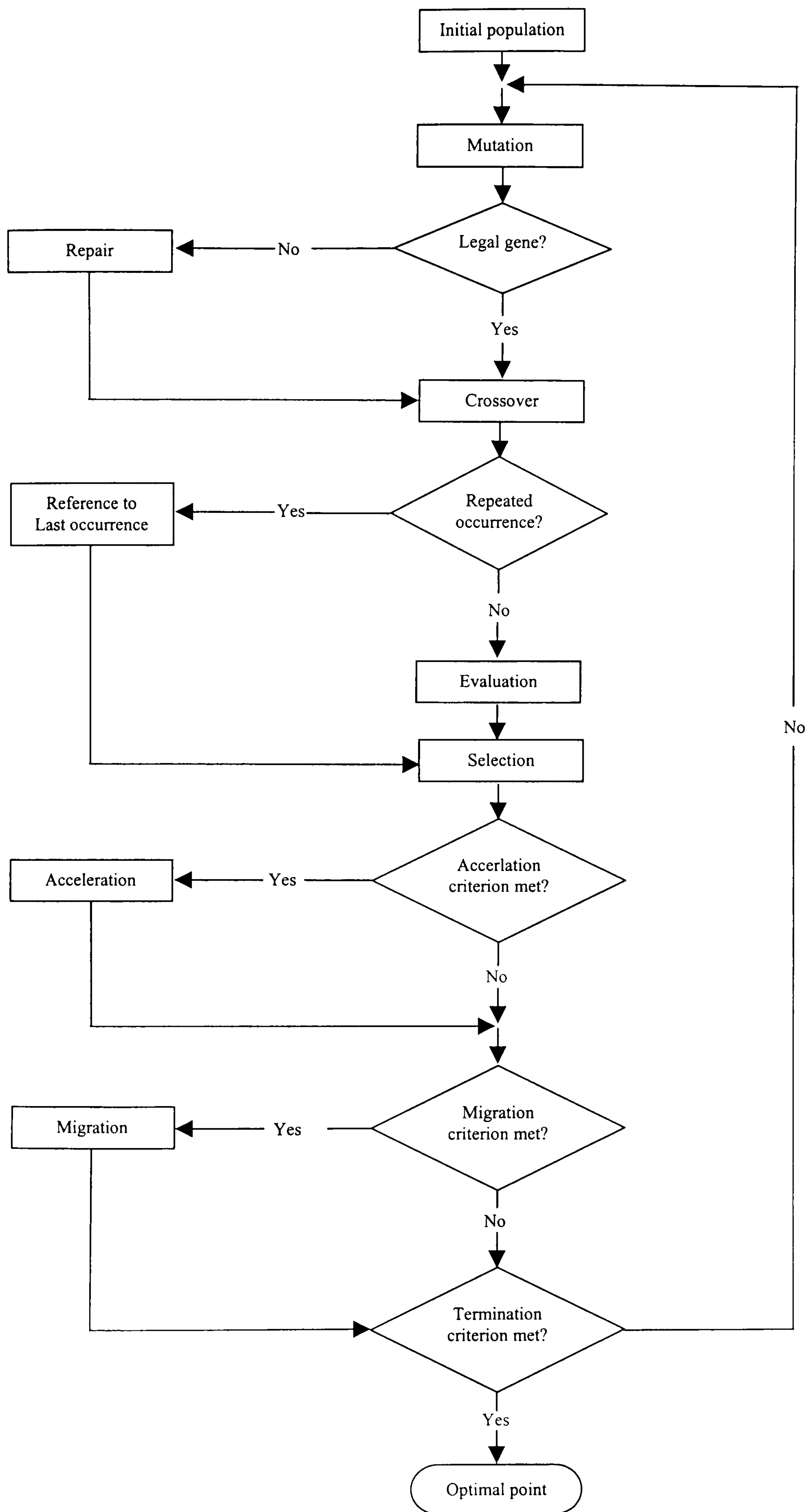


Fig. 2.6.1 A flowchart of the dvHDE method.

2.7 Convergence of the dvHDE method

Since the dvHDE method has been developed based on the DE method, the algorithm control parameters are the same as that for the DE, with some additional control parameters in the acceleration and migration operations. Some practical advice is given in [78] for tuning the control parameters of the DE method to get the best performance out of it. As an initial set up, the population size is suggested to be 10 times the number of parameters to be identified, and the mutation and crossover factors are set at 0.8 and 0.5 respectively. Adjusting the mutation factor to be slightly low or higher than 0.8 and/or increasing the population size should solve the problem if mis-convergence is experienced. There is always conflict between convergence speed and robustness. By increasing the population size and/or lowering the mutation factor, the possibility of convergence is higher, but with the expense of a longer run time. The DE has a fast convergence rate, as a result the population diversity rapidly descends, and this increases the probability of obtaining a local minimum. Using a larger population size may solve this drawback, however, by doing so, more computational time is required to evaluate the objective function. The problem can be crucial if the DE method is to be used in real world applications where much CPU time is required for solving differential equations. This was certainly the case when the DE method was used to solve parameter estimation problems of nonlinear systems, for examples the identification of hysteretic systems in [80,81], optimal control problem of a bioprocess in [83,84], and automotive application such as in [1]. A method that is able to find the global optimum point with the least computational time is therefore an advantage and is preferable. The proposed method in this work has been shown to be capable of achieving this by reducing the number of objective

function evaluations, and improving convergence speed while ensuring high population diversity thereby avoiding obtaining local minima and mis-convergence problems.

The performance of the dvHDE, like the DE method, is therefore affected by the values chosen for the control parameters; population size, n_i , mutation factor, μ_m , crossover factor, μ_c , and the scheme selected. At present, there are several variants of the DE method [75,78], the two particular strategies used throughout this work are the ‘DE/rand/1/bin’ and ‘DE/best/1/bin’ schemes. The ‘DE/rand/1/bin’ is an abbreviation for the strategy in which the DE uses a pair of chromosomes to generate the perturbation, $({}_gX_{i_1,j} - {}_gX_{i_2,j})$, that is added to the randomly selected third chromosome, ${}_gX_{i_3,j}$, in the mutation operation (equation 2.2), and the mutated ${}_g^mX_{i,j}$ and the current parent ${}_gX_{i,j}$ chromosomes are chosen using binomial distribution, in the crossover operation (equation 2.3). The perturbation may be calculated from more than one pair of chromosomes, and the crossover operation may employ other distribution such as exponential distribution. The ‘DE/best/1/bin’ denotes a similar strategy to the ‘DE/rand/1/bin’, the only difference is that the third chromosome ${}_gX_{i_3,j}$ in mutation operation is the best chromosome found from previous generation.

Generally, the convergence speed and robustness of the search process are influenced by values of μ_m and μ_c used. Most of the times, the control parameters, n_i , μ_m and μ_c are held constant during the search process. Randomly generated mutation factor ${}_g\mu_m = \Delta[1,1]$ and ${}_g\mu_m = \Delta[n_i, n_j]$, have also been experimented, where for generation

g^{th} , $\Delta[1,1]$ generates a random number between (0, 1), and $\Delta[n_i, n_j]$ generates a matrix of random numbers between (0, 1) sized $n_i \times n_j$. The ‘best’ values of μ_m and μ_c depend both on the population size and the characteristic of the objective function. The values are usually obtained by ‘trial-and-error’. The optimum values usually reflect the best compromise between speed of convergence and robustness. Population size and crossover factor control the number of offsprings that will potentially replace their parents. While, how many and the differences amongst the offsprings depend on the current population and mutation factor. The performance of the dvHDE method also depends on the performance of the acceleration and/or migration operations; when and how acceleration and migration are carried out. In the next section, three examples are considered to demonstrate the use and performance of the dvHDE method.

2.8 Numerical examples

The performance of the newly developed dvHDE method is best illustrated by applying the algorithm to a number of applications. First in this section, a mass-spring-damper single degree of freedom (1DOF) mechanical problem and two practical optimisations in engineering designs are considered as illustrative examples. The section aims to validate as well as to demonstrate the use of the proposed dvHDE method. The investigations here also explicitly shows where possible problems such as premature termination caused by local minima and mis-convergence may occur during the search for an optimum solution, and how the dvHDE overcomes them. The

estimation results are compared against those obtained from different numerical methods studied by other researchers.

The three optimisation examples are nonlinear programming problems, and they are chosen to illustrate different possible situations, that is where a different type of the parameters is involved; continuous, integer and a mixture of integer, discrete and continuous variables. The first problem is the parameter estimation of a simple mechanical model, whose parameters are all continuous variables. The second problem considers the design of a gear train. This example represents an optimisation problem that contains integer variables. The third problem is the design of a coil spring. In this example the optimisation is a nonlinear programming problem that involves a mixture of integer, discrete and continuous variables. This example represents a large fraction of the engineering design optimisation problems. In general, when discussing nonlinear programming, the variables of the objective function are usually assumed to be continuous. It is however very common that optimisation problems also involve discrete and/or integer variables. For example, the design variables are commonly discrete because the available values are limited to a set of commercially standard sizes; the sizes and numbers of bolts or rivets needed to fix a structure, the thickness of a steel plate, the diameter of a copper tube are some examples. After investigating the three problems, some conclusions are drawn regarding the performance and use of the dvHDE method.

2.8.1 Single degree of freedom vehicle problem

A large number of vibration applications employ a combination of the elements such as masses, springs, and dampers used in modelling the system under investigation. In this example, a vehicle dynamics problem is considered. For acceleration, braking, and most turning analyses, a vehicle, which consists of many components, can be represented as one lumped mass located at its centre of gravity with appropriate mass, m , and inertia properties. The vehicle suspension may be considered to have the total or equivalent stiffness and damping properties of k and d respectively. At its simplest, one may then consider a vehicle as a 1DOF system consisting of a body of mass m , spring, k and damper, d , as shown in Fig. 2.8.1.

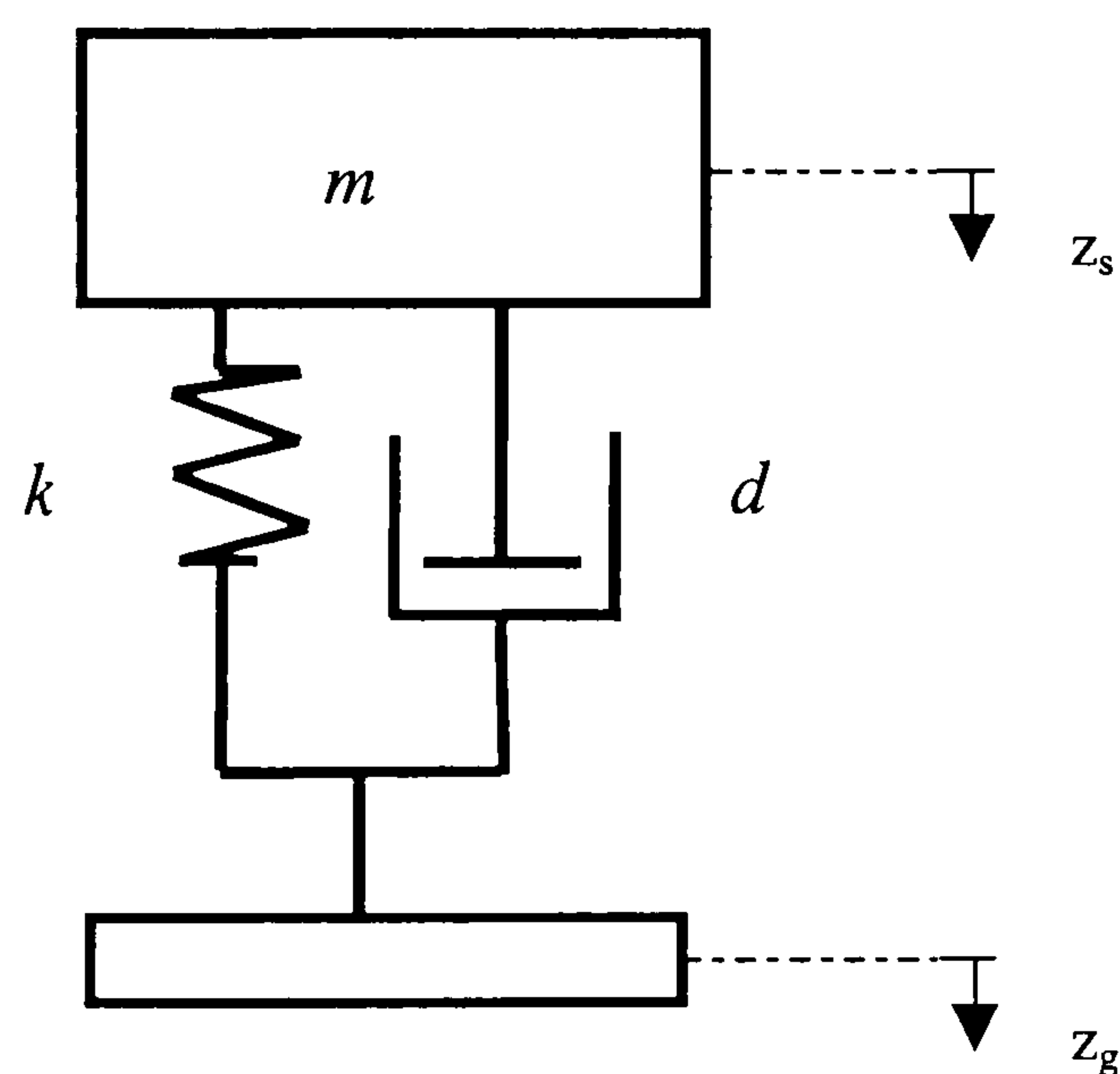


Fig. 2.8.1 Single degree of freedom (1DOF) vehicle model.

The aim here is to estimate the values of m , k and d from the vehicle response to a given input. The estimation problem thus contains three parameters, which are all continuous variables. Suppose, in this illustrative case, that it is sufficient to know the

mass accurately up to 0.01 kg, the spring stiffness to 1 N/m, and damping coefficient to 0.1 Ns/m. The parameter ranges for the numerical search to look for are given as follows; mass in the range of 0-500 kg, stiffness in the range of 0-100 kN/m, and the damping coefficient in the range of 0-5000 Ns/m. These a-priori information, which will be used by the optimisation algorithm, are therefore as follows; see also equation (2.1),

The parameter lower bound, $\theta_{\min,j} = [0, 0, 0]$

The parameter upper bound, $\theta_{\max,j} = [5 \times 10^2, 100 \times 10^3, 5 \times 10^3]$

The parameter precision required, $\theta_{tol,j} = [0.01, 1.0, 0.10]$

$j = 1, 2, \dots, n_j$, and $n_j = 3$

For the dvHDE algorithm, a gene is thus an index of the mass, spring stiffness and the damping coefficient, and a chromosome comprises of four genes; the fourth one being its identifier.

In this example, the vehicle response to be used in the parameter estimation is obtained by a computer simulation. The simulated system's equation of motion is given by

$$m \ddot{z}_s(t) = -d(\dot{z}_s(t) - \dot{z}_g(t)) - k(z_s(t) - z_g(t)) \quad (2.9)$$

with $m = 200$ kg, $k = 15$ kN/m, and $d = 700$ Ns/m. $z_g(t)$ is the system input and $z_s(t)$, $\dot{z}_s(t)$, $\ddot{z}_s(t)$ are the displacement, velocity, and acceleration of the body respectively. The performances of the DE and dvHDE methods are discussed below.

A typical plot of the fitness values belonging to the best chromosomes at each generation by the DE and dvHDE method is depicted in Fig. 2.8.2. To improve the clarity, the values are plotted against log of generation. It is evident from the plot that the DE method has a fast convergence rate, bearing in mind that fast convergence might lead to a high probability of getting local minima. The best fitness value quickly decreases at the early generations and steadily approaches the optimum point at the later generations. It is seen also that there are several consecutive generations that the DE fails to update the best fitness values or only makes small reductions, resulting in the DE's curve appearing 'discontinuing'. The 'discontinuous' characteristic of the DE is dealt with in the dvHDE method. The acceleration operation successively decreases the best fitness value from generation to generation. The fitness values for the dvHDE method thus appears 'smoother' and converges even faster to the optimum point.

As generation progress, both in the DE and dvHDE methods, the population eventually clusters closely at one point in the search space, that is the point corresponding to the best chromosome. A plot of the population diversity in Fig.2.8.3 illustrates this fact. Considering the DE method, it can be seen that the population diversity rapidly decreases. After about 45 generation, the diversity measure reduces to and remains zero. This reflects two facts; first, the whole population is close to the best individual. Second, the mutation and crossover operations are not capable of improving the population diversity. This means if the algorithm is trapped in a local minimum point, the DE could fail to reach the optimum point.

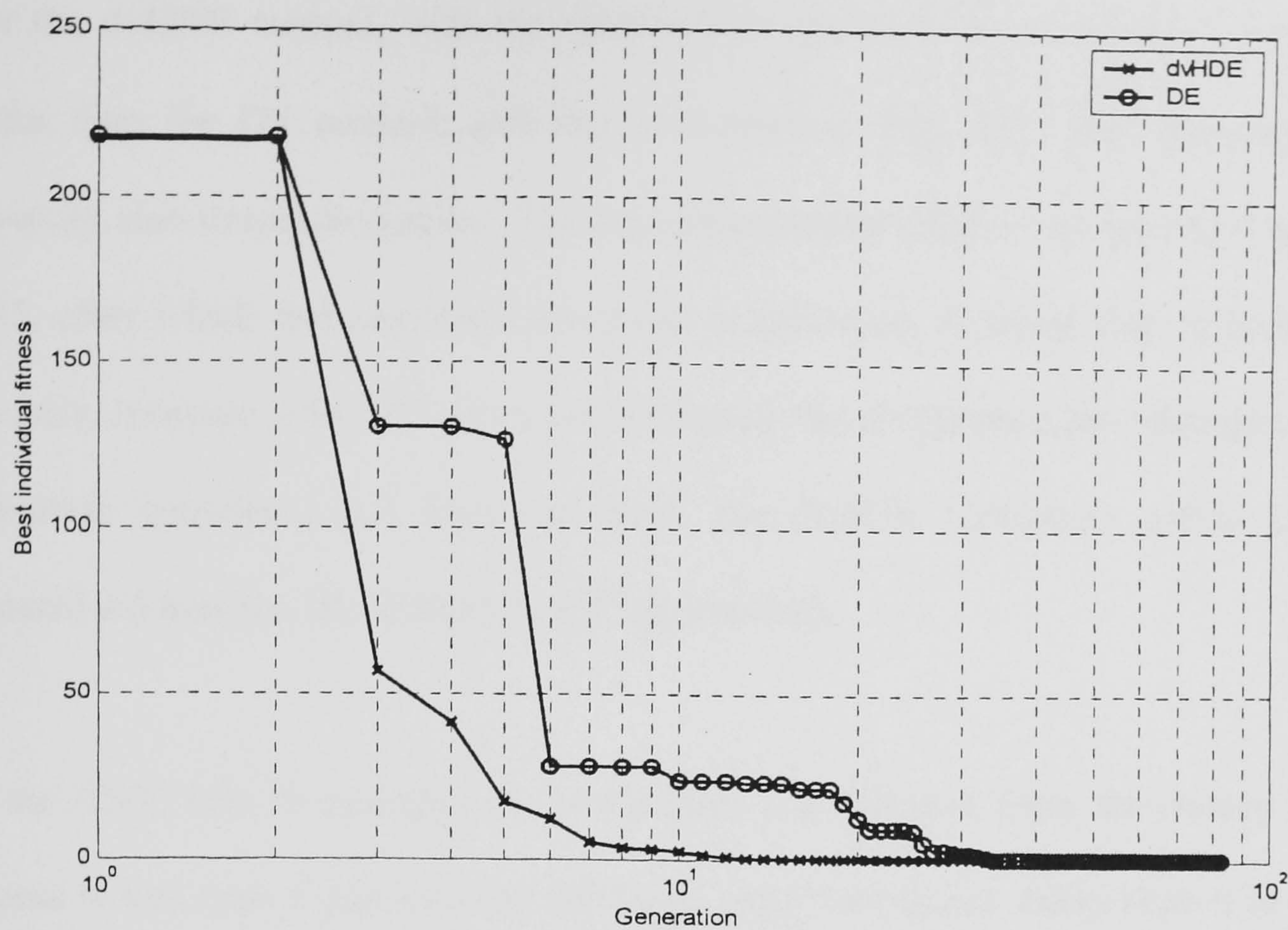


Fig. 2.8.2 A typical plot of fitness values belonging to the best individual at each generation.

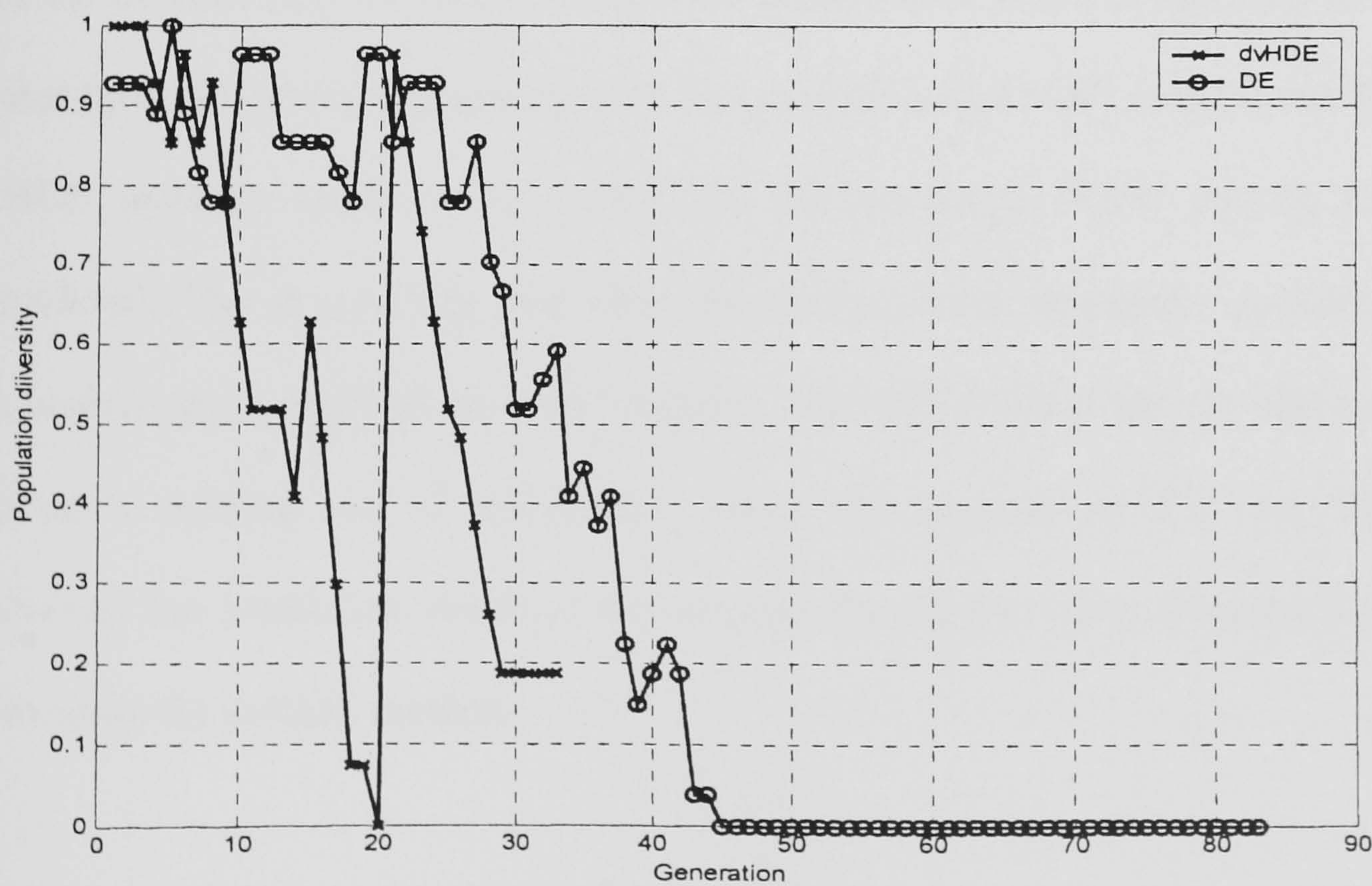


Fig. 2.8.3 A typical plot of population diversity.

For the dvHDE method, with the acceleration operation incorporated, it converges faster than the DE method, and this is evident in Fig. 2.8.3 that the population diversity also descends quicker. The population reduces below the specified value of 0.05, after which the migration operation is activated. A sharp rise of population diversity measure at the 20th generation indicates that a migration has taken place. The evolution continues, and from the plots, the dvHDE terminates earlier (at 34th generation) than the DE method (at 83rd generation).

In the 1DOF vehicle example, the acceleration is performed when the change in best fitness is less than 5 percent. The DS loops only 5 iterations every time it is called. From 20 independent runs, the average number of evaluations is 598 for the dvHDE and 617 for the DE method. In terms of numbers, the dvHDE reduces the number of evaluation operation by a small percentage, 3.1%. However, it must be remembered that the dvHDE has included the migration operation, a mean of ensuring obtaining global optimum point. Comparing with the gradient-based (GB) method, the DE and dvHDE methods tends to be slower for the simulated 1DOF vehicle problem considered. This is generally true when the correct model structure is assumed, and the search space contains no local minima. The GB method has an advantage of efficiently making use of information about the gradient of the cost function. However, the simulation result in this section has shown some advantages gained from using the dvHDE method.

2.8.2 Gear train design problem

The second example to be examined is a gear design application. The design consists of four spur gears as depicted in Fig 2.8.4; the driver, the follower and two intermediate gears. The driver gear, which has x_1 teeth, provides the input torque to the two intermediate gears, whose axis is shared, and the torque is in turn transferred to the follower, whose number of teeth is x_4 . The gear connected to the driver has x_3 teeth, and the one connecting to the follower has x_2 teeth.

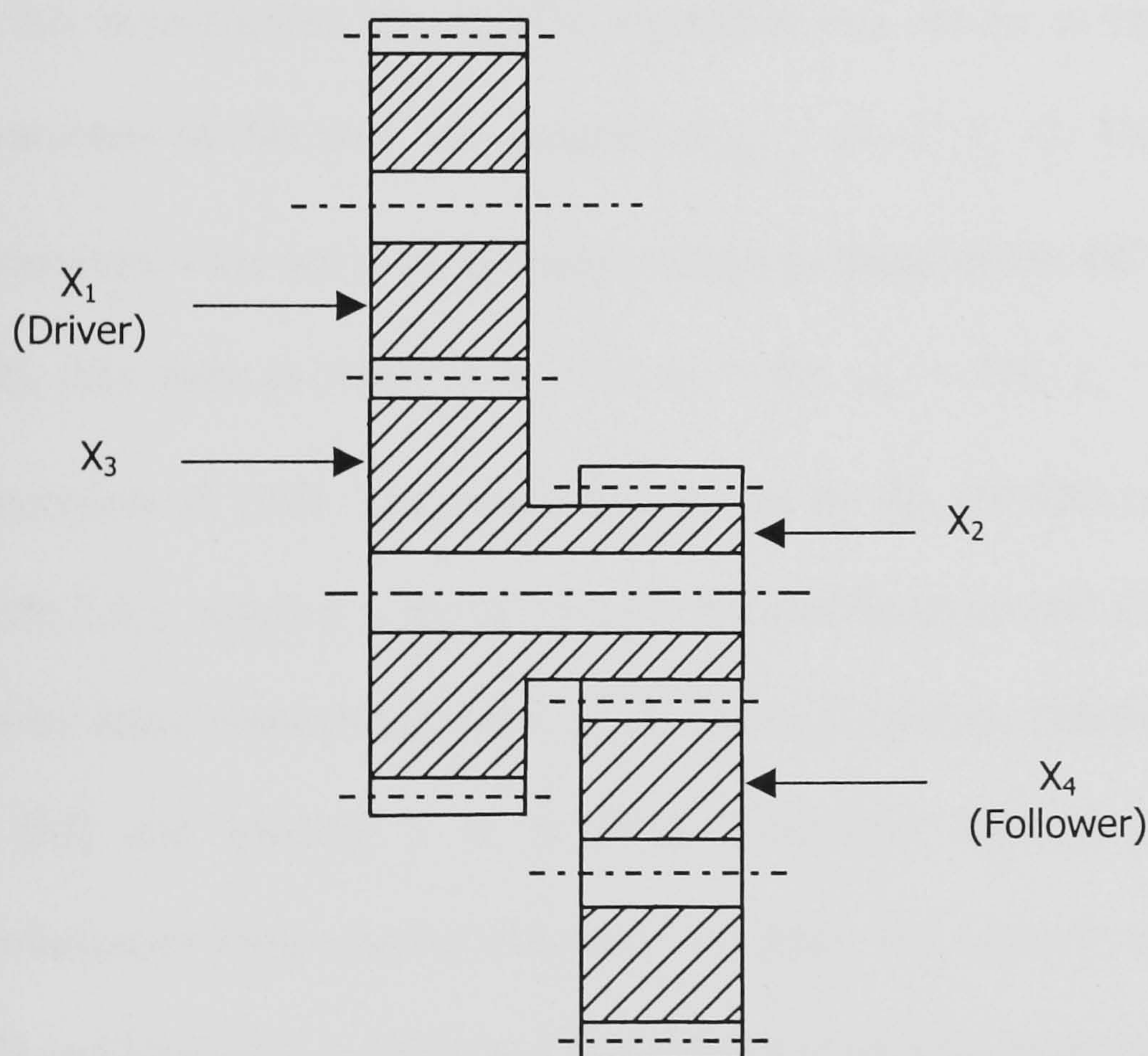


Fig. 2.8.4 Gear train.

The objective of the design is to maximise the torque delivered to the follower for a given torque on the driver gear by choosing the number of gear teeth, x_1 , x_2 , x_3 and x_4 .

The maximisation of the output torque can be formulated as a parameter estimation problem, where the involved parameters, which are all integers, are the number of

gear teeth. The search for the best four parameters is treated as the minimisation of equation (2.10) subjected to the constraints given in equation (2.11).

$$J(X) = \left(\frac{1}{6.931} - \frac{x_1 x_2}{x_3 x_4} \right)^2 \quad (2.10)$$

subject to

$$12 \leq x_j \leq 60, \quad j = 1, 2, \dots, 4 \quad (2.11)$$

In this investigation the dvHDE algorithm was set-up as follows. Since all of the parameters in this case are integers, $\theta_{tol,j} = [1, 1, 1, 1]$. The other dvHDE control parameters were set to be the same values as those of the DE method investigated in [90], they were as follows; $n_j = 4$, $n_i = 40$, $\mu_m = 0.8$, $\mu_c = 0.7$, with a maximum generation of 1000. The estimation results by the dvHDE method are presented in Table 2.8.1, together with the results obtained from several different methods carried out by other researchers; three gradient-based methods (Method 1 in [91], Method 2 in [92] and Method 3 in [93,94]), Simulated Annealing (Method 4 in [95]), Evolutionary Programming (Method 5 in [96]), two Genetic Algorithms (Method 6 in [97] and Method 7 in [98]), and Differential Evolution (Method 8 in [90]).

Result	Method1	Method2	Method3	Method4	Method5	Method6	Method7	Method8	dvHDE
x_1	18	14	19	30	30	19	19	16	16
x_2	22	29	16	15	15	16	16	19	19
x_3	45	47	42	52	52	49	43	43	43
x_4	60	59	50	60	60	43	49	49	49
$J(X)$	5.7×10^{-10}	4.5×10^{-6}	0.23×10^{-6}	2.36×10^{-9}	2.36×10^{-9}	2.7×10^{-12}	2.7×10^{-12}	2.7×10^{-12}	2.7×10^{-12}
Gear Ratio	0.1467	0.1464	0.1448	0.1442	0.1442	0.1443	0.1443	0.1443	0.1443
Error [%]	1.6500	1.1700	0.3340	0.0340	0.0340	0.0011	0.0011	0.0011	0.0011

Table 2.8.1 Gear train design: Parameter estimation results by different methods.

Method1: Branch and bound using sequential quadratic programming [91].

Method2: Integer-discrete-continuous non-linear programming [92].

Method3: Sequential linearisation algorithm [93,94].

Method4: Simulated Annealing [95].

Method5: Evolutionary Programming [96].

Method6: Modified genetic algorithm [97].

Method7: Meta-genetic algorithm [98].

Method8: Differential Evolution [90].

From Table 2.8.1, the dvHDE method was able to find the optimum point with $J(X) = 2.7 \times 10^{-12}$, which is the same value as that of the Genetic Algorithms (Method 6 and Method 7), and the DE method (Method 8). The method has also identified the three other solutions with the equal final value of $J(X)$ when;

$$x_1 = 16, x_2 = 19, x_3 = 49, x_4 = 43,$$

$$x_1 = 19, x_2 = 16, x_3 = 49, x_4 = 43,$$

$$x_1 = 19, x_2 = 16, x_3 = 43, x_4 = 49,$$

which are just permutation on the solution given.

Without incorporation of the acceleration and migration operations only 62% of the runs reached the optimum point. When the solution did not converge, the search showed signs of experiencing two problems. First, there existed several local minima, see Fig. 2.8.5 (a) and also Fig 2.8.6 (a). The search evolved quickly and found a local minimum. Recalling that a high convergence speed can result in a greater possibility of obtaining a local minimum. Without any other mechanisms, the search was trapped in a local valley, after which the mutation and crossover were only able to update local chromosomes within the valley. The population diversity eventually reduced to zero or very close to zero, which meant that the population clustered around the best chromosome. The search carried on until the maximum generation criterion was met. The situation of hitting a local minimum point is demonstrated Fig. 2.8.5(a). The accumulated number of function evaluations increased with generation number. From the plot, approximately after generation 140th, where the population diversity was very close to zero, the accumulated number of function evaluations no longer increased.

(b)

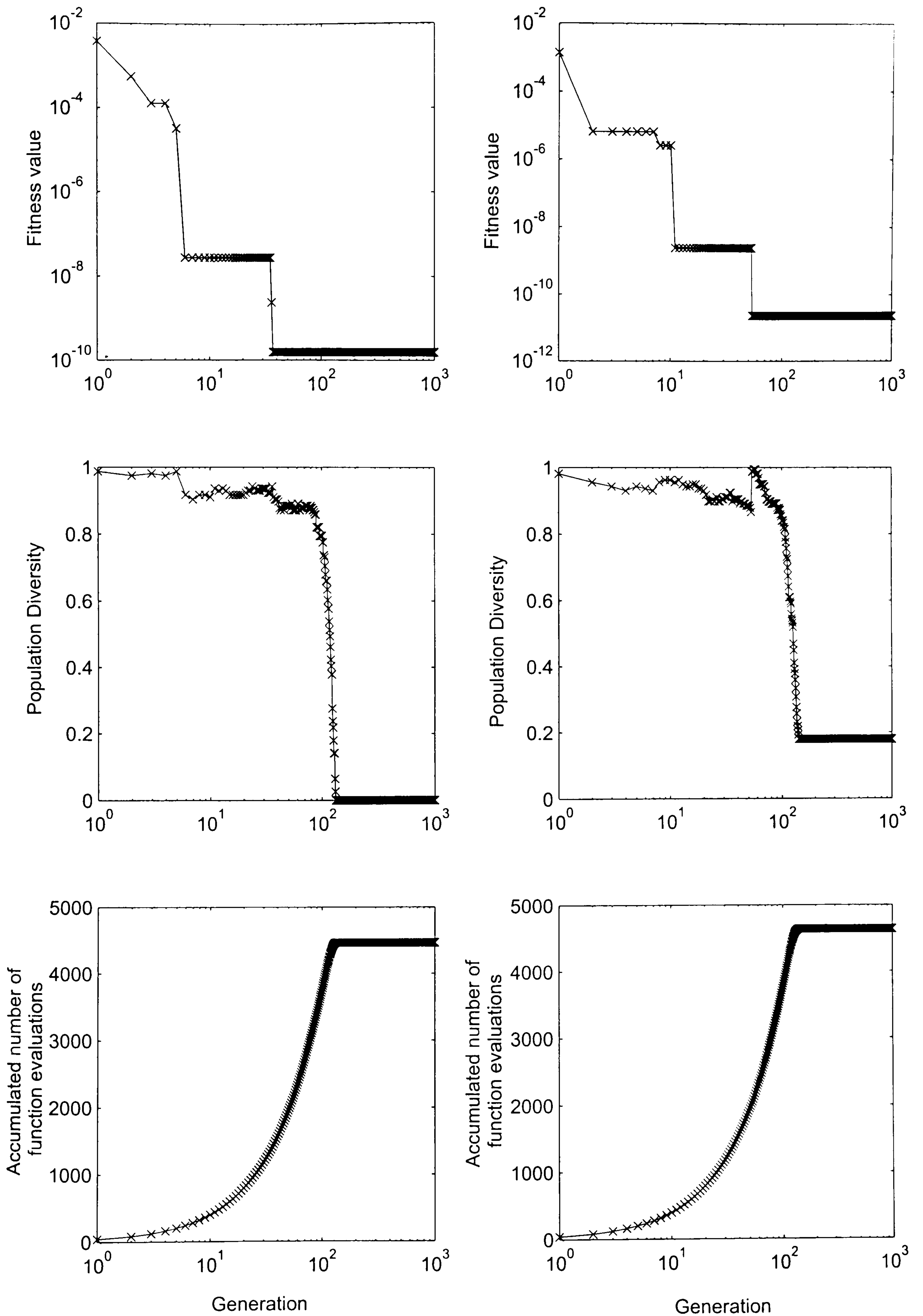


Fig 2.8.5 Plots of two encountered convergence problems; (a) premature due to local minima, (b) stagnation.

This was because, with a low diversity, there was no ‘new’ offsprings generated. All of the offsprings that could possibly be produced by that present population were already once assigned fitness values. The dvHDE recognised their occurrence, thus performed no further evaluations. Fig. 2.8.5(a) therefore illustrated three important symptoms of premature convergence problems; first, the population had converged to a local minimum point of a multi-modal objective function, second, the population had lost its diversity, and lastly, the search proceeded slowly or did not proceed at all.

The second problem encountered was stagnation. The problem of stagnation differs from a premature convergence problem, and is illustrated in Fig. 2.8.5(b). In this particular example, the dvHDE method, like the DE method, showed a sign of possibly experiencing stagnation. The indication of encountering stagnation was that the search had not converged to a local optimum or any other point. It however no longer updated the best-fit chromosome, while the population was still remaining diverse. As can be seen from the plot of population diversity, it unchanged over a number of generations, before the search terminated as the maximum generation criterion was reached. This suggested the mutation and crossover operation could not produce offsprings that were better than their corresponding parents in the selection operation. Thus the search no longer evolved. When the search stagnated, two situations may be distinguished. Firstly, the search was still able to produce various chromosomes, but none of them succeeded to replace their corresponding parents. Secondly, the number of offsprings that could possibly be produced was fixed, and they were all less fit than their parents. This second situation was encountered more often than the first.

Observing the accumulated number of chromosome fitness evaluations, together with the fitness value and population diversity, helped diagnose the two convergence problems described. In the first situation the accumulated number of fitness evaluations increased as the generation evolved. While in the second case, the accumulated number of evaluations remained the same, since all the possible offsprings that would be produced by the current population had already been evaluated once. When not converged, the dvHDE method still managed to find a solution with $J(X)$ equal to or lower than 1.36×10^{-9} , even better than results given in the gradient-based methods (Method 1-Method 3), Simulated Annealing (Method 4) and Evolutionary Programming (Method 5), see Table 2.8.1.

In an attempt to overcome local minima and stagnation problems, the first solution was to increase the population size from $n_i = 40$ to $n_i = 80$. Increasing the population size reduces the possibilities of premature termination and stagnation. The result was that 90% of the runs found the optimum point, with the average number of function evaluations of 27813, see Fig. 2.8.6(c). Increasing the population size has considerably improved the mis-convergence problem. When not converging to an optimum point, the search failed to escape from the local minimum at $J(X) = 2.3 \times 10^{-11}$. The possible risk of entering local minima valleys or encountering stagnation was reduced by a larger population pool. The inherent drawback was more chromosomes were needed to be assigned fitness values. Though the dvHDE avoided re-evaluations of ‘already-occurred once’ chromosomes, however because of a larger population size a greater number of ‘newly occurred’ chromosomes needed be assigned fitness through out, from initialisation to termination.

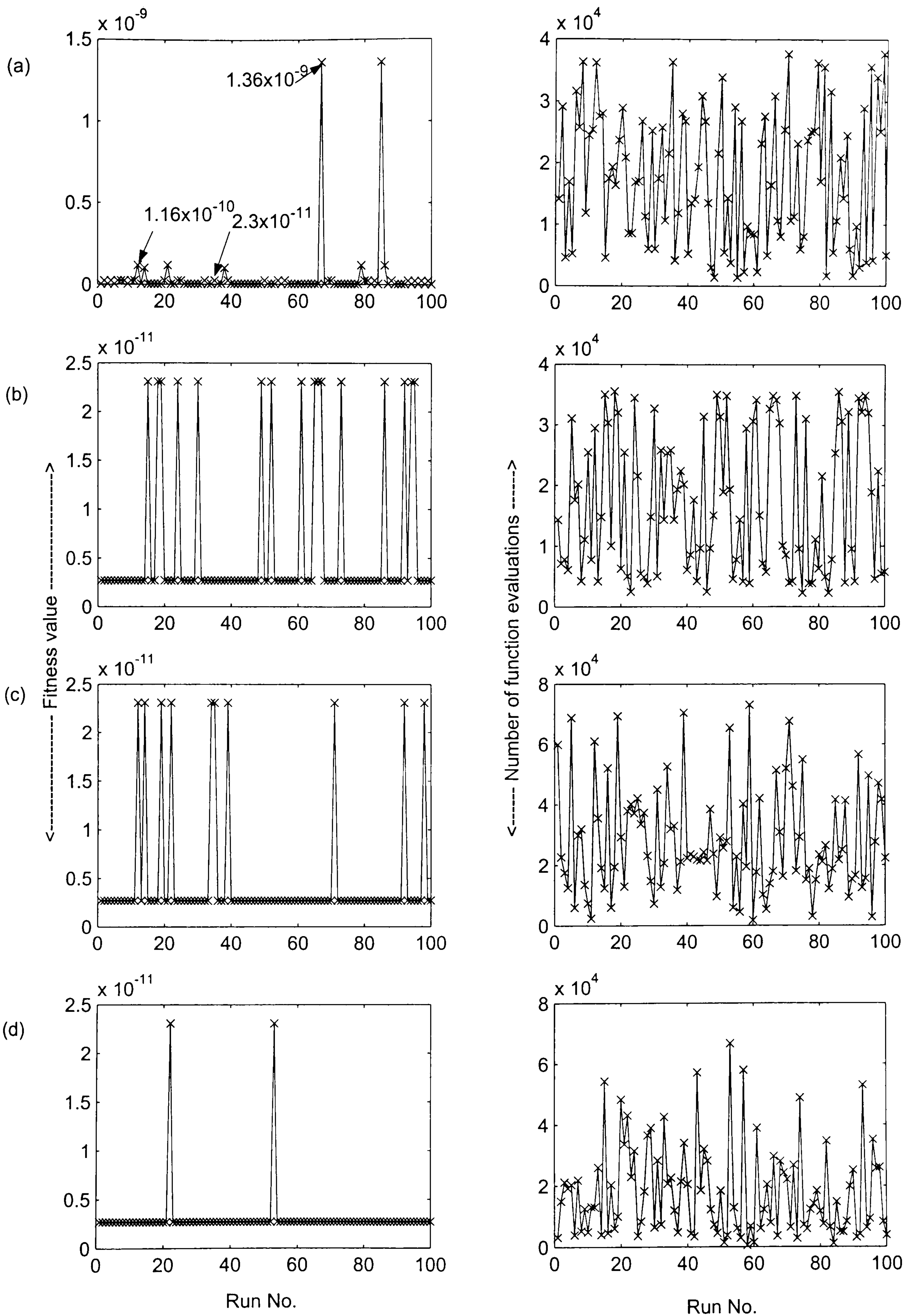


Fig 2.8.6 Plots of results from 100 runs when (a) $n_i = 40$, (b) $n_i = 40$, with Migration, (c) $n_i = 80$, (d) $n_i = 80$, with Migration.

Without increasing the population size, the second solution was to incorporate the migration operation. Injecting a newly produced population by the migration operation would provide a mean of possible escape from local minima and the stagnation problem. The newly produced population can be thought of as if they were additional chromosomes, however they were only introduced and replace the current population when a specified condition was met. In this particular design problem, the strategy was to perform the migration operation when one or both of the following conditions was met.

The first condition was to perform the migration when the population diversity was less 0.2 (that is the population diversity tolerance $\epsilon_1 = 0.2$, see equation (2.7) and (2.8)), with the gene diversity tolerance, ϵ_2 , set at 0.05.

The second condition was initiated when the search showed a sign of encountering stagnation, this was determined by monitoring both the fitness value and the number of function evaluations.

With the population size, $n_i = 40$ and incorporation of the migration, the result was 84% of the runs found the optimum point, with an average number of function evaluations of 17053, see Fig. 2.8.6(b). Though the number of successful runs was 6% less than that when $n_i = 80$, however, in term of the number of function evaluations, employing the migration operation without increase the population size required 38.7% less number of function evaluations. Thus, the ‘gain’ from time saving was more than the ‘loss’ in the reduced success rate to find the global minimum, as justified by the percentage figures. The lower number of successful runs may be

partially due to the population size of 40, which was too small that even inclusion of the migration operation could not produce a success rate of more than 90%. When combined with an increase in population size (from $n_i = 40$ to $n_i = 80$) with the migration operation, the dvHDE was able to find the optimum point 98% of the runs, with an average number of function evaluations of 17789, see Fig. 3.8.6(d). The number of function evaluations was approximately 36% less than that when only the population size was increased. Therefore in this case, the ‘gain’ from time saving was 36% and the success rate was also increased by 8% compared to when $n_i = 80$ and without the migration. With the migration incorporated, setting $n_i = 80$ further increased the number of successful runs by 14% from when $n_i = 80$, however at an expense of 2.7% more function evaluations being carried out. The demand on the algorithm to consistently find the global optimum thus conflicts with the time it required to do so. The ‘best’ set-up for the algorithm is again a compromise between robustness and convergence speed. However all in all, by increasing the population size and incorporating the migration has been proved to be successful and the best solution to the problems of premature termination and the stagnation for the gear design problem considered.

2.8.3 Coil spring design problem

The second parameter estimation problem considered was a non-linear mechanical engineering design optimisation of a coil spring. The problem was first studied by Sandgren [91] and later by other researchers [90,98-100]. The task is to minimise the volume of spring steel wire used to manufacture the spring. The problem involved

estimation of three design parameters; one integer, one discrete and one continuous. The integer variable is the number of spring coils and is denoted by x_1 . The discrete variable, which is the spring wire diameter chosen from a set of available standard size provided in Table 2.8.2, is denoted by x_2 . The continuous variable denoted by x_3 is the outside diameter of the coil spring, see also Fig. 2.8.7.

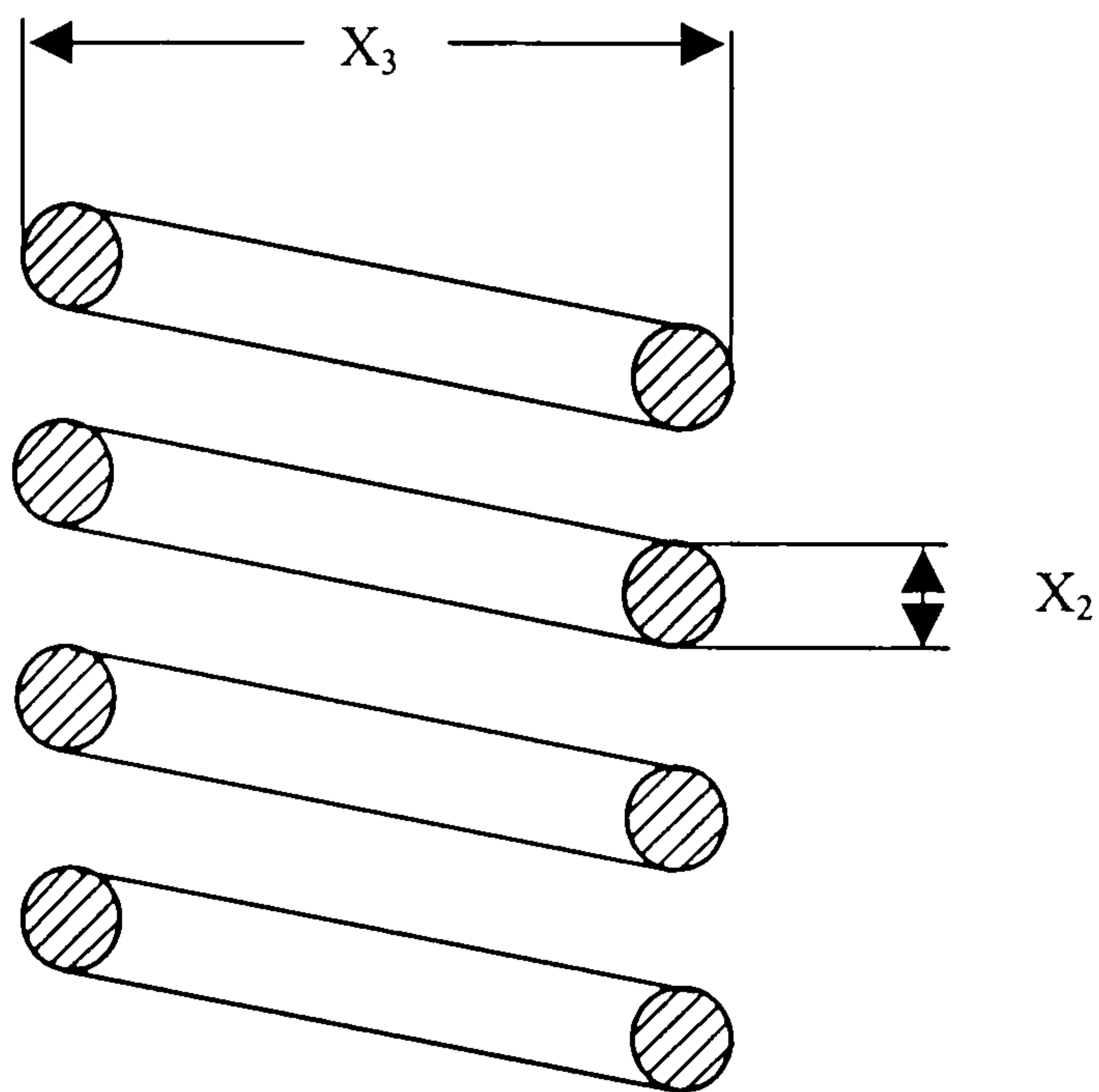


Fig. 2.8.7 Coil spring.

Available wire diameters [inch]					
0.009	0.0095	0.0104	0.0118	0.0128	0.0132
0.014	0.015	0.0162	0.0173	0.018	0.020
0.023	0.025	0.028	0.032	0.035	0.041
0.047	0.054	0.063	0.072	0.080	0.092
0.105	0.120	0.135	0.148	0.162	0.177
0.192	0.207	0.225	0.244	0.263	0.283
0.307	0.331	0.362	0.394	0.4375	0.500

Table 2.8.2 Available wire diameters, x_2 .

The parameter estimation was formulated with chromosome fitness function given in equation (2.12), subjected to eight constraints $g_1(X) - g_8(X)$.

$$J(X) = \frac{\pi^2 x_2 x_3^2 (x_1 + 2)}{4} \quad (2.12)$$

$$g_1(X) = \frac{8C_f F_{\max} x_2}{\pi x_3^3} - S \leq 0$$

$$g_2(X) = l_f - l_{\max} \leq 0$$

$$g_3(X) = d_{\min} - x_3 \leq 0$$

$$g_4(X) = x_2 - D_{\max} \leq 0$$

$$g_5(X) = 3.0 - \frac{x_2}{x_3} \leq 0$$

$$g_6(X) = \sigma_p - \sigma_{p \max} \leq 0$$

$$g_7(X) = \sigma_p + \frac{F_{\max} - F_p}{K} + 1.05(x_1 + 2)x_3 - l_f \leq 0$$

$$g_8(X) = \sigma_p - \frac{F_{\max} - F_p}{K} \leq 0$$

where

$$C_f = \frac{4(x_2 / x_3) - 1}{4(x_2 / x_3) - 4} + \frac{0.615 x_3}{x_2} \leq 0$$

$$K = \frac{G x_3^4}{8x_1 x_2^3}$$

$$\sigma_p = \frac{F_p}{K}$$

$$l_f = \frac{F_{\max}}{K} + 1.05(x_1 + 2)x_3$$

The design constant values are given below.

- a) The maximum working load, $F_{max} = 1000.0$ lb (453.5970 kg.)
- b) The allowable maximum shear stress, $S = 189000.0$ psi (1302.9660×10^6 N/m²)
- c) The maximum free length, $l_{max} = 14.0$ inch (0.3556 m.)
- d) The minimum wire diameter, $d_{min} = 0.2$ inch (0.0051 m.)
- e) The maximum outside diameter of the spring, $D_{max} = 3.0$ inch (0.0762 m.)
- f) The pre-load compression force, $F_p = 300.0$ lb (136.0791 kg.)
- g) The maximum allowable deflection under pre-load, $\sigma_{p\max} = 6.0$ inch (0.1524 m.)
- h) The deflection from pre-load position to maximum load position,
 $\sigma_w = 1.25$ inch (0.0318 m.)
- i) The shear modulus of the material, $G = 11.5 \times 10^6$.

Two settings were investigated and compared; setting 1: $n_i = 40$, $\mu_m = 0.9$, $\mu_c = 0.9$, and setting 2: $n_i = 40$, $\mu_m = 0.7$, $\mu_c = 0.8$. These values were set to be the same as those of the DE method investigated in [90], so that direct comparison between the performance of the dvHDE and DE method can be made. For both setting, $\theta_{tol,j} = [1, 1 \times 10^{-12}, 1]$, and the maximum generation set at 1000. The performance of the dvHDE is discussed in three cases; case 1: without the acceleration and migration operations, case 2: with the migration operation only, case 3: with both the acceleration and migration operations.

For the first setting, the result from 100 independent runs by the dvHDE method without incorporating the acceleration and migration operations (case 1) is presented in Fig.2.8.8(a).

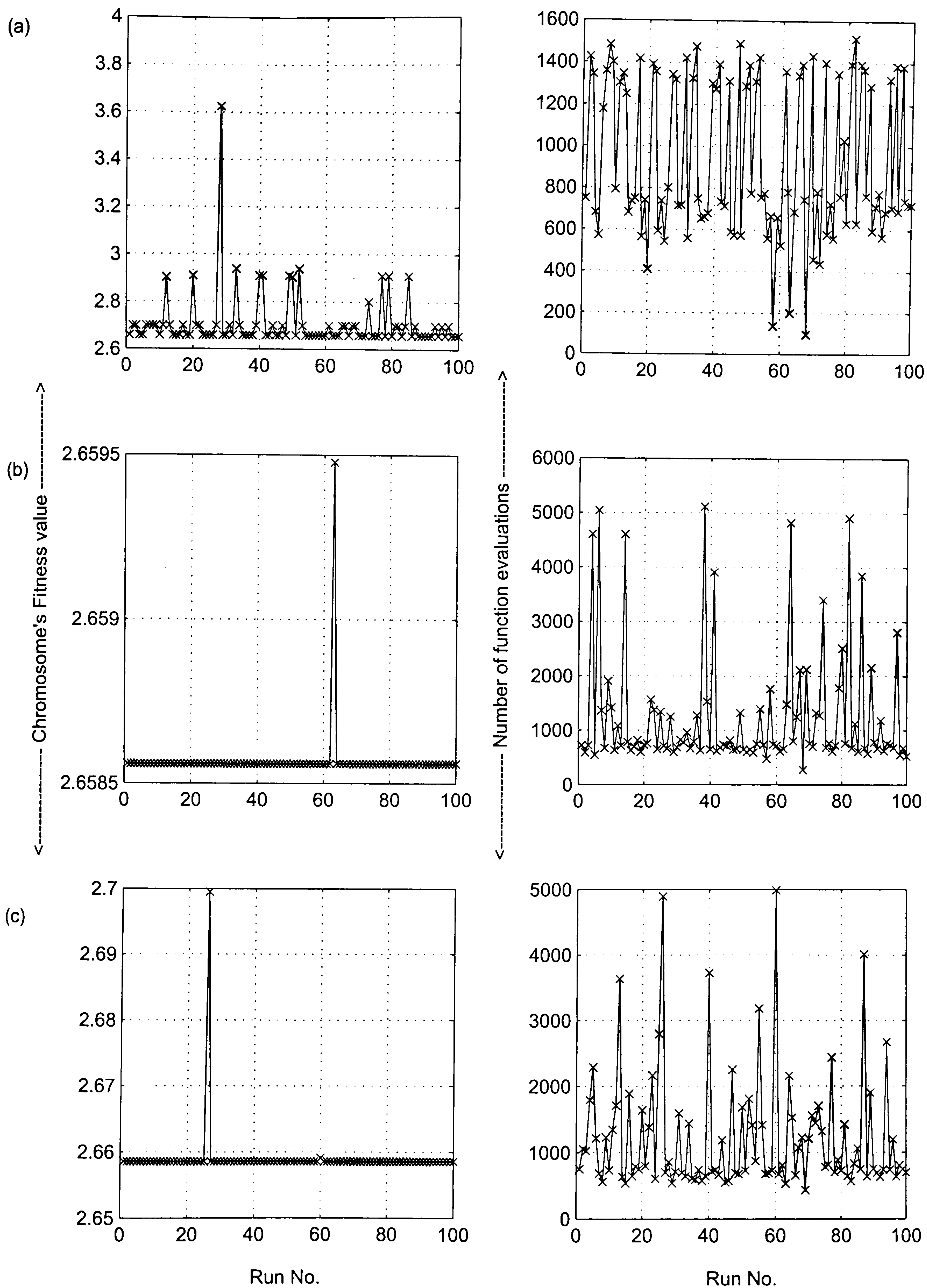


Fig 2.8.8 Plots of results for setting 1; $\mu_m=0.9$, $\mu_c=0.9$ when (a) without Migration and Acceleration, (b) with Migration, (c) with Migration and Acceleration.

Though the number of function evaluations required was low, with an average of 935.5, the search was able to find the optimum point only with a low percentage of success 57%. As can be seen from Fig. 2.8.8(a) the search terminated at, not less than four, different places. This, together with the observation of population diversity, suggested that the search encountered several local minima. The search did not show signs of experiencing a stagnation problem. When including the migration operation (case 2), the results were improved dramatically, with a success of 99%, see Fig. 2.8.8(b). The plot of the corresponding number of function evaluations shows the difference when compared with that of case 1. The average number of function evaluations increased to 1247.1. When the search encountered a local minimum, the dvHDE method attempted to overcome the problem using the migration operation. As a consequence, the number of function evaluations increased. The success of the migration operation depended on the newly injected population, the dvHDE control parameters, n_i, μ_m, μ_c , and the migration strategy including the choice of the desired tolerance for the population diversity and the gene diversity, ε_1 and ε_2 respectively. In terms of percentage of success and number of function evaluations required, the dvHDE method has been shown to be better than those results reported in literatures [90,91,98,100], the results are given in Table 2.8.3.

The result found by the DE method reported in [90] required 26000 function evaluations when a maximum generation set at 650 generations. To improve the result further, an acceleration operation was also included in the dvHDE (case 3), the percentage of success was then 99%, with number of function evaluations slightly reduced to 1223, see also Fig 2.8.8(c).

	Method1	Method2	Method3	Method4	dvHDE
x1	10	9	9	9	9
x2	0.283	0.283	0.283	0.283	0.283
x3	1.1807	1.2287	1.2274	1.2230	1.2230
g_1	54309	415.9690	550.9930	1008.8114	1008.8114
g_2	8.8187	8.9207	8.9264	8.9456	8.9456
g_3	0.0830	0.0830	0.0830	0.0830	0.0830
g_4	1.8193	1.7713	1.7726	1.7770	1.7770
g_5	1.1723	1.3417	1.3371	1.3217	1.3217
g_6	5.4643	5.4568	5.4585	5.4643	5.4643
g_7	0.0	0.0	0.0	2.6758×10^{-16}	0.0
g_8	0.0	0.0174	0.0134	5.0751×10^{-16}	5.9130×10^{-13}
$J(X)$	2.7995	2.6709	2.6681	2.6586	2.6586
Successful run	100.0%	95.4%	95.3%	95.0%	99.0%

Table 2.8.3 Optimal solutions for coil spring problem.

Method1: Branch and bound using sequential quadratic programming [91].

Method2: Genetic algorithm [100].

Method3: Meta-genetic algorithm [98].

Method4: Differential Evolution [90].

For the second setting, where $\mu_m = 0.7$, $\mu_c = 0.8$, the results from 100 runs are presented in Fig. 2.8.9. Similar to setting 1, setting 2 without the acceleration and migration operations encountered several local minima. The convergence speed slightly increased with the average number of function evaluations being 926.5, with a decrease of robustness from 57% to 47%, see also Fig. 2.8.9(a).

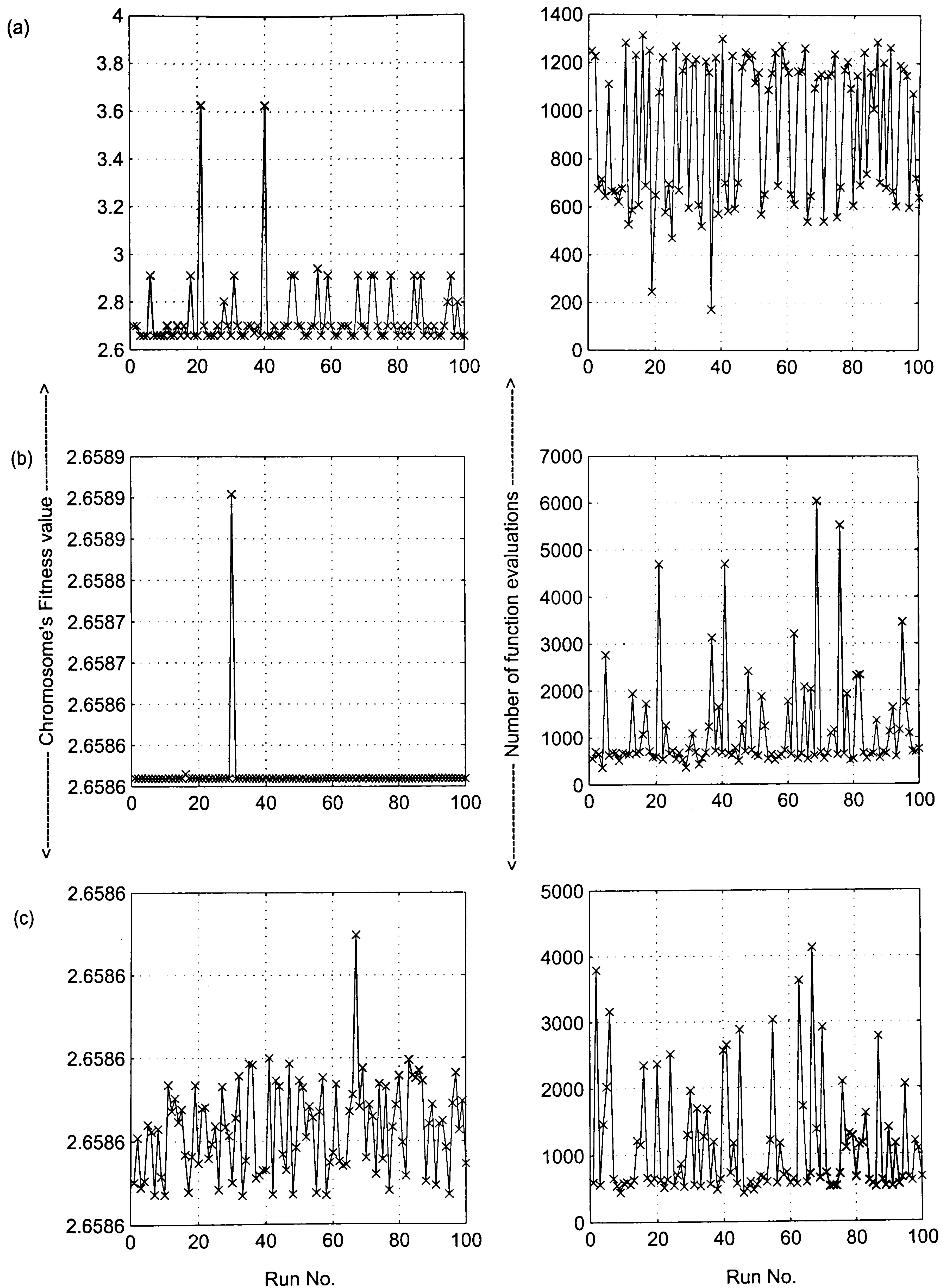


Fig 2.8.9 Plots of results for setting 2; $\mu_m = 0.7$, $\mu_c = 0.8$ when (a) without Migration and Acceleration, (b) with Migration, (c) with Migration and Acceleration.

Inclusion of the migration (case 2) again showed significant improvement in the consistence of finding the optimum point with 99 % success, for an increase in the number of function evaluations to 1148.1, see Fig 2.8.9(b). When the acceleration and migration were combined (case 3), from Fig. 2.8.9(c) it can be seen that the search found the optimum point every time, and with slightly less number of function evaluations, 1127.3, when compared with case 2. The result reported for the DE method in [90] with $n_i = 40$, $\mu_m = 0.7$, $\mu_c = 0.8$, and a maximum of 200 generations, required 8000 function evaluations. Again, the dvHDE method has shown to be significantly better than the other methods reported in Table 2.8.3. The success of the dvHDE method for this particular problem came from the inclusion of the migration operation, with further small improvement when the acceleration was included. It was possible to perform the migration operation several times while keeping the number of function evaluations to a minimum. The success of the migration and acceleration operations not only depends on the current population and the algorithm control parameters, but also on the strategy, which decides when and how the acceleration and migrations should be carried out.

2.9 Conclusions

In this chapter, the motivations to the development of the proposed discrete variable Hybrid Differential Evolution (dvHDE) method were discussed. The important motivation came from unsatisfactory results involving problems of obtaining local minima and time consuming due to slow convergence speed when applying the conventional gradient-based (GB) and Differential Evolution (DE) methods. After

giving a brief description of the DE, on which the dvHDE is based, the functional mechanisms of the dvHDE algorithm were explained in details. The dvHDE method was then applied to three test problems to validate and illustrate its potential use. The investigation on the test problems had provided opportunities to examine the performance of the dvHDE method when including and excluding the two additional operations, the acceleration and migration. The results have been discussed and compared with other techniques including the GB and DE methods.

The first problem, which was a simulation study on a 1DOF mass-spring-damper system, has illustrated advantages of the dvHDE over the DE method. Firstly, the undesirable ‘discontinuous’ descending characteristics of the fitness value in the DE method has been dealt with in the dvHDE method using the acceleration operation. The acceleration operation successively decreased the best fitness value from generation to generation resulting in the dvHDE method converged to the global optimum in a smaller number of generations when compared to the DE method. Secondly, the DE and dvHDE methods both were shown to have fast convergence rates, as a consequence, the population diversity rapidly reduced and eventually reached zero. For the DE method, this means the algorithm has a high possibility of failing to reach the global optimum, if it is trapped in a local minimum valley. This problem was accounted for in the dvHDE method using the migration operation, which improved the population diversity by injecting a newly generated population, thus preventing premature termination and mis-convergence problems.

In the second test problem, which was a gear train design, the result obtained by the dvHDE method was compared to other techniques reported in the literatures. The

dvHDE found the same result as those of the Genetic Algorithms and the DE method, and a better result than that of the Simulated Annealing, the Evolutionary Programming and the gradient-based (GB) methods. The investigation identified and explicitly illustrated two possible convergence problems; premature termination due to local minima and mis-convergence due to stagnation. The investigation also illustrated the performance of the dvHDE method when attempting to overcome the two convergence problems experienced. The results from 100 independent runs have shown that without the acceleration and migration operations the percentage of successful run was low (62%). The incorporation of the migration has improved the result to 84% success rate. Alternatively, by increasing the population size the number of successful runs was improved to 90%, however, in an expense of greater number of function evaluations. A combination of increase in population size and incorporation of the migration was shown to be the best solution to the problems of premature termination and stagnation for the gear design. The result was 98% success rate, with 36% less number of function evaluations compared to when increasing the population size alone.

The third test problems, which was a coil spring designs involving a mixture of integer, discrete and continuous variables, was shown to experience several local minimum points. Without the acceleration and migration, the success rate was 57%, and was improved dramatically by the migration operation to 99%, with a small increase in number of function evaluations. The best solution has been shown to be a combination of the migration and acceleration operations, where the result was 99% success with less number of function evaluations when compared to incorporating the migration alone. Comparing to other techniques, the dvHDE method obtained a better

result than that of the GAs and the gradient-based method, and found the same result as that of the DE method, however, with a higher percentage of success and approximately 86% less number of function evaluations.

The investigation on the three test problems has illustrated potential use of the proposed dvHDE method. The algorithm's performance depends not only on its control parameters (the population size, mutation and crossover factors) but also the strategies of how and when the acceleration and migration are carried out. There is always a conflict between the convergence speed and the consistency of finding the global optimum point when setting up for the algorithm's control parameters and incorporating the acceleration and migration operations. As demonstrated in the second test problem, increase in population size may help solving local optima and stagnation problems, however, with an inherent drawback of greater number of function evaluations. The acceleration operation can be performed to improve the convergence speed, and the migration operation can be incorporated to prevent premature termination and mis-convergence. The two additional operations thus act as a trade-off for each other. However, performing either or both of the two operations usually requires a greater number of function evaluations, thus more time consumed. The dvHDE method has also attempted to reduce the number of function evaluations both during the recombination and acceleration and migration operations by avoiding re-evaluations of the chromosomes that have already been appeared. For the three test problems, which are small sized problems, the number of repeated chromosomes is significant. The benefit gained from checking and avoiding re-evaluation is however only appreciated when the number of repeated chromosome, which depends on the problem size, is significant.

The values of the control parameters and the strategies for the acceleration and migration are problem dependent. The best choice for these should reflect the best compromise between algorithm robustness and convergence speed. Initially the dvHDE method should be attempted without the acceleration and migration operations. When the search shows signs of encountering premature termination and mis-convergence problems, the dvHDE with migration operation should be attempted first. A combination of the acceleration and migration operations can then be included if the problems of premature and/or mis-convergence is encountered, and convergence speed is of important.

The performance of the dvHDE method has shown to be the best for the limited number of problems considered here, and needs to be extended in future work to further examine its performance when applied to different applications and different problem size.

CHAPTER 3 PARAMETER ESTIMATION OF AUTOMOTIVE DAMPERS

The work of the previous chapter has shown the possible benefits of the dvHDE method when applied to engineering problems. The numerical examples considered were non-linear programming problems, however, they did not involve experimentation. In this chapter, it is aimed to further investigate the performance of the dvHDE method when used on the parameter estimation of a more complicated problem from experimental data. This investigation involves both modelling and experimentation work on an automotive damper for a wheeled passenger car.

Automotive dampers are one of the most important components in modern vehicle suspension systems [1-5,101-106]. They are known to exhibit complex behaviour when tested. The knowledge about their dynamic behaviour when the vehicle is travelling over a particular terrain under certain circumstances is important for both safety and comfort, since vehicle engineers can implement the information to the design and/or service of the dampers themselves or the vehicle suspension as a whole. For example, during the life cycle of an automotive damper the performance of the damper declines due to wear and loss of its working oil. This is a very slow process and therefore is usually not noticed over time by the driver who gets accustomed to the changing behaviour of the vehicle. It is useful if such faults in the dampers can be detected and then corrected accordingly for safety and comfort reasons. A damper model can be used in the fault detection of a damper, since changes in its behaviour are usually linked to variation of the model parameters. Therefore in such case the damper model and its parameters prove necessary.

In this chapter system identification and parameter estimation for an automotive damper is exercised. The task here is to choose the best model within the model structure selected for the test damper. This involves selecting a model structure and then estimating the model parameters. The identified damper model can be used to study the damper behaviour and its performance when it is incorporated into the vehicle suspension system or full vehicle, within the frequency range of interest for vehicle ride, which is between 0.5-30 Hz.

The chapter also aims to investigate how different numerical search methods perform when the optimisation problem becomes more complex. The problem's complexity may increase, not only, due to the number of model parameters but also in other respect, such as the numerical solution to the dynamic behaviour of the model. The estimation results and performance of the proposed dvHDE method is compared with those of the standard gradient-based (GB), the Downhill-Simplex (DS) and the original DE methods.

This chapter first introduces some important points relating to the dynamic behaviour of a vehicle damper. Various tests on the damper have been carried out in order to gain a better understanding of its dynamic behaviour, and also to collect the data for the estimation procedure. The experimental set-up is described and the experimental data are presented and discussed in section 3.2. Analysis of the data suggested a direction for the modelling of the damper within the frequency range of interest 0.5-30 Hz. A review on the modelling of automotive dampers is given in section 3.3. The modelling approach chosen for the damper is to use a model whose parameters

represent its important physical properties. Five different models were proposed and tested using the experimental data. The model parameter estimation is a single objective optimisation problem, in which the difference between the measured and modelled damper forces is to be minimised. The qualities of fit for the five models are analysed in details. The performance of the four numerical search methods for each model considered are then discussed and compared. Several important conclusions are drawn regarding the performance of the dvHDE method and the modelling of automotive dampers in section 3.5.

3.1 Introduction

A primary function of the vehicle suspension system is to isolate the vehicle structure, occupants and goods so far as is practicable from shock loading and vibration due to irregularities of the road surface. It must do this without impairing the stability, steering or general handling qualities of the vehicle. Dampers are used, together with suspension springs and linkages, as parts of the suspension system to achieve those requirements. A damper is often called a shock absorber, however the component does not absorb shock but dampens or dissipates the stored energy in the suspension system. When suspension springs are deflected, they absorb energy, and the dampers will dissipate some of the energy from the springs. Dampers also give the suspension engineer another mean of tailoring the suspension movement for a given application. They can be designed to resist suspension movement, bounce, body roll, brake dive or acceleration squat.

Hydraulic dampers are commonly used in modern vehicle suspension systems. As mentioned, the demand on their characteristics is not only to achieve the desired ride characteristics, but also must satisfy the function of keeping good tyre-to-road contact, which is essential for handling and safety. The characteristic of the damper is therefore not a simple linear one, and it deserved more detailed consideration in the modelling and estimation stages. In fact, the reaction force generated by a damper is a function of several variables, including the displacement, velocity, acceleration and frequency of the motion imposed. A sketch layout of typical hydraulic dampers in an automotive application is provided in Fig. 3.1.1; (a) gas-pressurised monotube, and (b) twin tube. Although the two are different in detailed construction, they consist of common parts such as a piston rod, rod seals, piston and valves. In addition to the compression and rebound chambers, the twin tube damper has a reserve chamber to accommodate volume changes in the compression chamber. As a consequence, the twin tube design requires a base valve controlling flow between the compression and reserve chambers. The twin tube damper could be designed to have the reserve chamber outside the damper unit, using a linked container. This would have a benefit of avoiding unwanted mixture of the working oil and gas, and also allow for better cooling. Alternative to having a reserve chamber, the monotube design uses a floating piston or a bag filled with a pressurised-gas to achieve compensation of pressure changes.

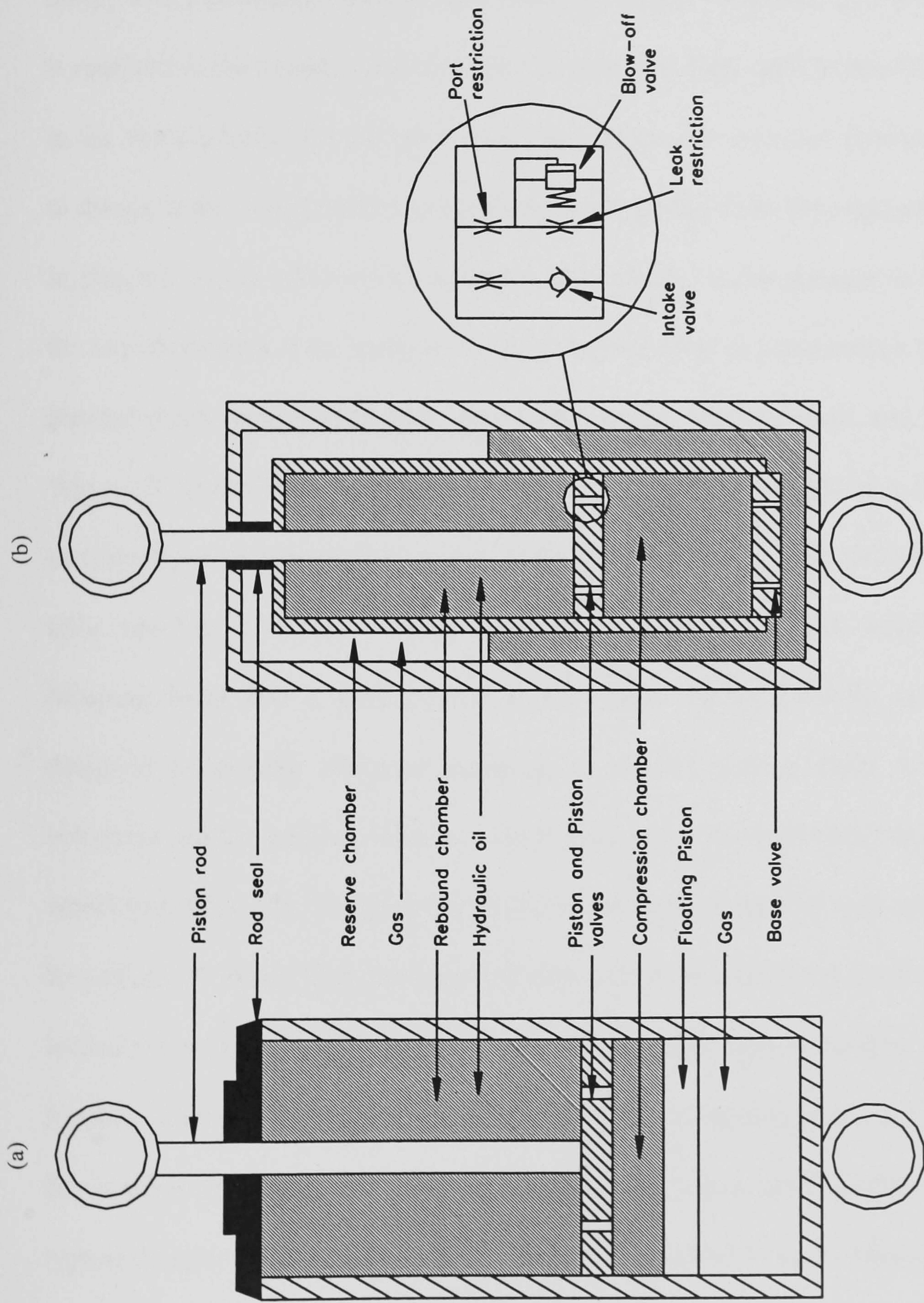


Fig. 3.1.1 A sketch layout of typical hydraulic automotive dampers; (a) Gas-pressurised monotube, and

(b) Twin tube, also showing a schematic valve assembly.

The dynamic behaviour of dampers used in vehicle applications is purposely designed to have different damping characteristics in compression (when the damper rod moves in to the damper body) and rebound (when the damper rod moves out of the damper body). For a passenger vehicle, as a wheel encounters a bump, only a small damping is required in the compression direction as damping force adds to the force transmitted to the vehicle body. On the other hand, damping in the rebound direction is desirable to dissipate the energy stored in the suspension springs from the encounter with bump. In general, this is achieved by choosing the 'valving' in the damper to be different in the two directions. Two types of valving may be used in combination to produce the desired characteristic [101,102]. One is an orifice control valve and the other is a 'blow-off' control valve. The orifice control valve may consist of a port restriction and intake valve, which allows flow of the working fluid in a predetermined direction only, see Fig. 3.1.1 (b). A port restriction is a large channel, which generates a damping force that is proportional to the square of the velocity. However, when designed to provide adequate damping to control sprung mass motions at low velocities, orifice control yields too much damping at the high velocities typical of the wheel-hop mode [3]. To solve this problem, the port restriction may also be followed by a blow-off valve. With the blow-off valve, the flow passage is blocked by a spring-loaded valve so that it prevents flow until a desired pressure is reached, at which point it opens allowing 'blow-off' with a damping force as shown in curve 2 of Fig. 3.1.2. By combining orifice and 'blow-off' controls in series and parallel arrangements, typical damper behaviour such as that indicated by curve 3 can be obtained.

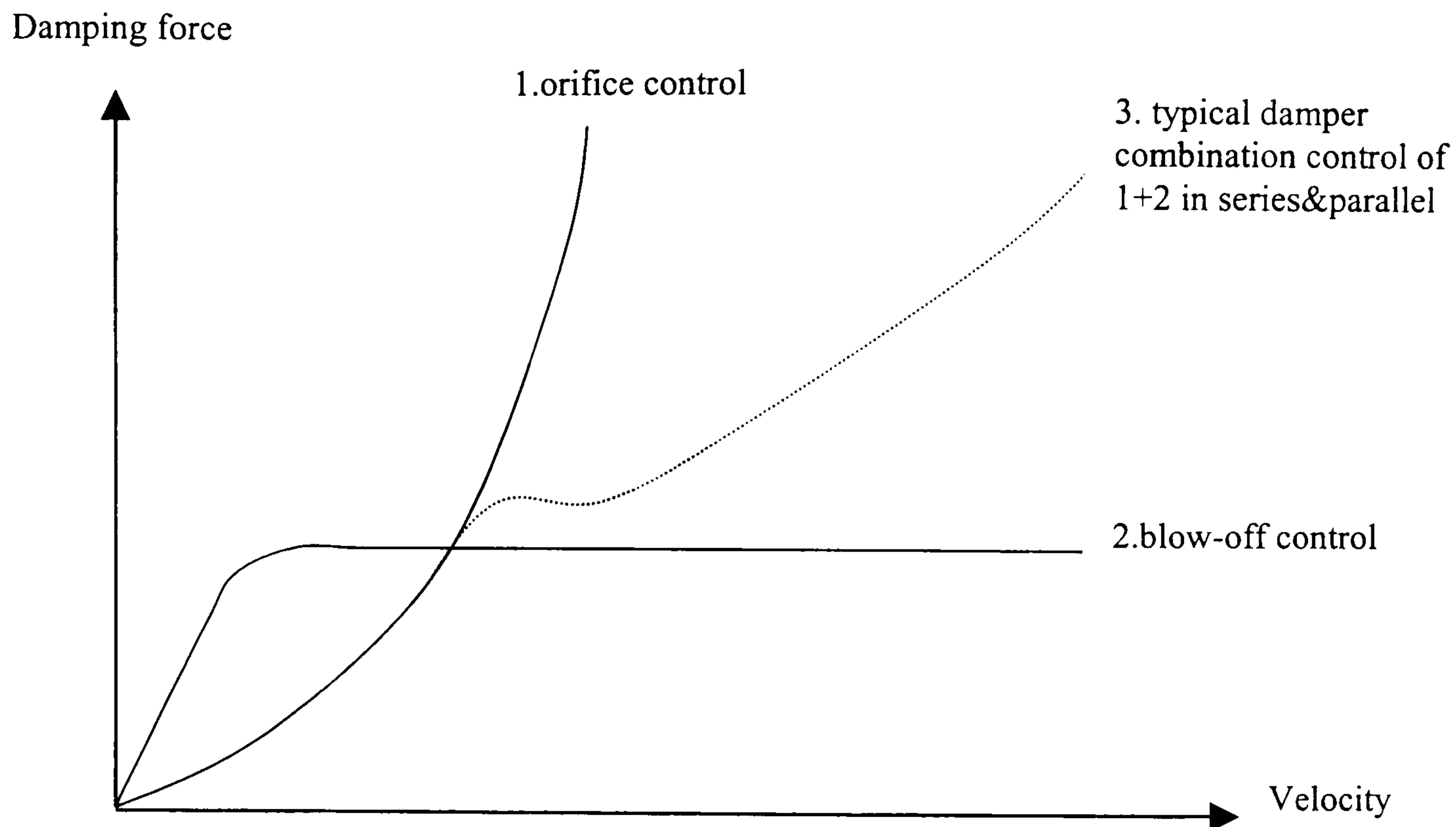


Fig. 3.1.2 Damping force-velocity characteristics of typical dampers [3].

Apart from nonlinear behaviour, there are other inherent features that are often present, to a smaller or greater extent. These are mainly in the form of dry friction and hysteresis. Friction is induced by relative motions such as that between the damper rod, the rod guide, piston bearing and the rod sealing. Hysteresis is caused by the compressibility of the hydraulic fluid, the compliance in the chamber walls and the rubber bushings and sudden changes in force due to vaporisation of the fluid or expansion of entrapped gasses. In addition, the dependency of the damper characteristics on the working fluid temperature, due to the kinetic energy when it is converted into heat through losses caused by the fluid flow through orifices, is also to be expected.

3.2 Damper Test

Various testing on the damper were performed to develop a better understanding of its dynamic behaviour helping in the modelling phase. Section 3.2.1 describes the test rig and the experimental set-up. The test was carried out in the Engineering Dynamics Centre at Royal Military College of Science (RMCS), Cranfield University. The experimental data and analysis of data are presented in section 3.2.2.

3.2.1 Experimental Set-up

The experimental apparatus was set up as shown in a photo in Fig. 3.2.1. A diagram of the test rig and instrumentation is also given in Fig 3.2.2. The surrounding frame incorporating the damper was made from substantial welded steel section. The four-corners of the frame were supported on rubber suspension blocks, which in turn were secured to the concrete floor. The test facility consisted of a hydraulic actuator and a servo-controller. The hydraulic actuator was located at the top of the frame, and was controlled by the Kelsey Instruments servo-controller. The actuator head has a rated capacity of 70 kN static force and 70 kN peak dynamic force. The exciter head comprised of several major components. They included a hydraulic cylinder of stroke 160 mm, six servo valves operating in parallel, a hydraulic filter, high pressure and exhaust accumulators, and a displacement transducer (LVDT) which provided a signal in the range of ± 10 V that was linearly proportional to the stroke of the actuator cylinder. The Kelsey servo-controller (K7500) was designed to enable the static and dynamic level of the actuator head to be controlled independently and simultaneously.

The types of feedback available included both force and displacement control. The displacement command input signal to the test rig could be square, tri-angular and sine waves with constant amplitude or frequency. An arbitrary generated input profile can also be fed through from a signal generators or a PC to the controller via the external-operating mode. This allowed investigation of the damper behaviour under excitation of different input profiles.

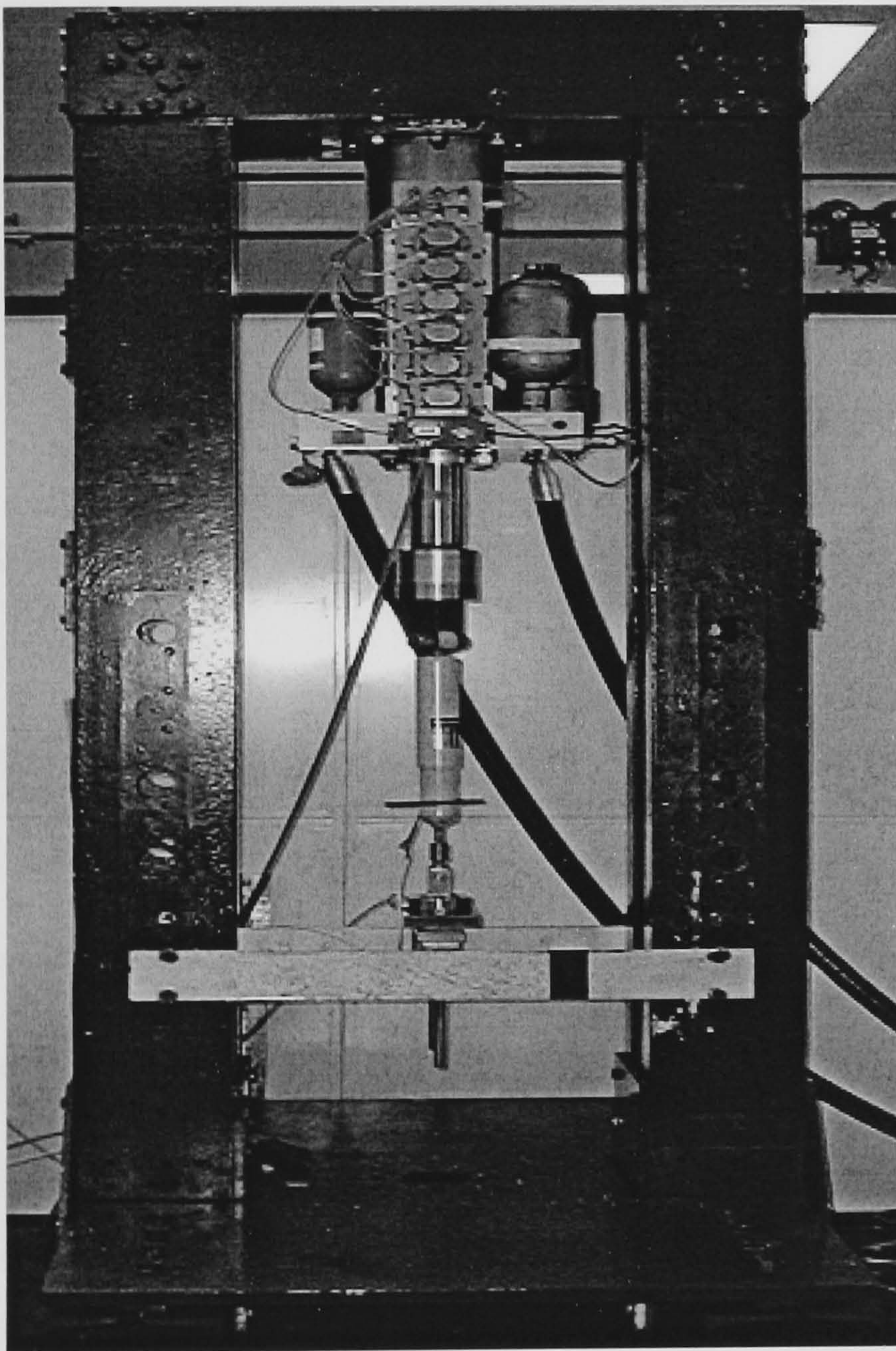


Fig. 3.2.1 Photo of the Damper Test Rig.

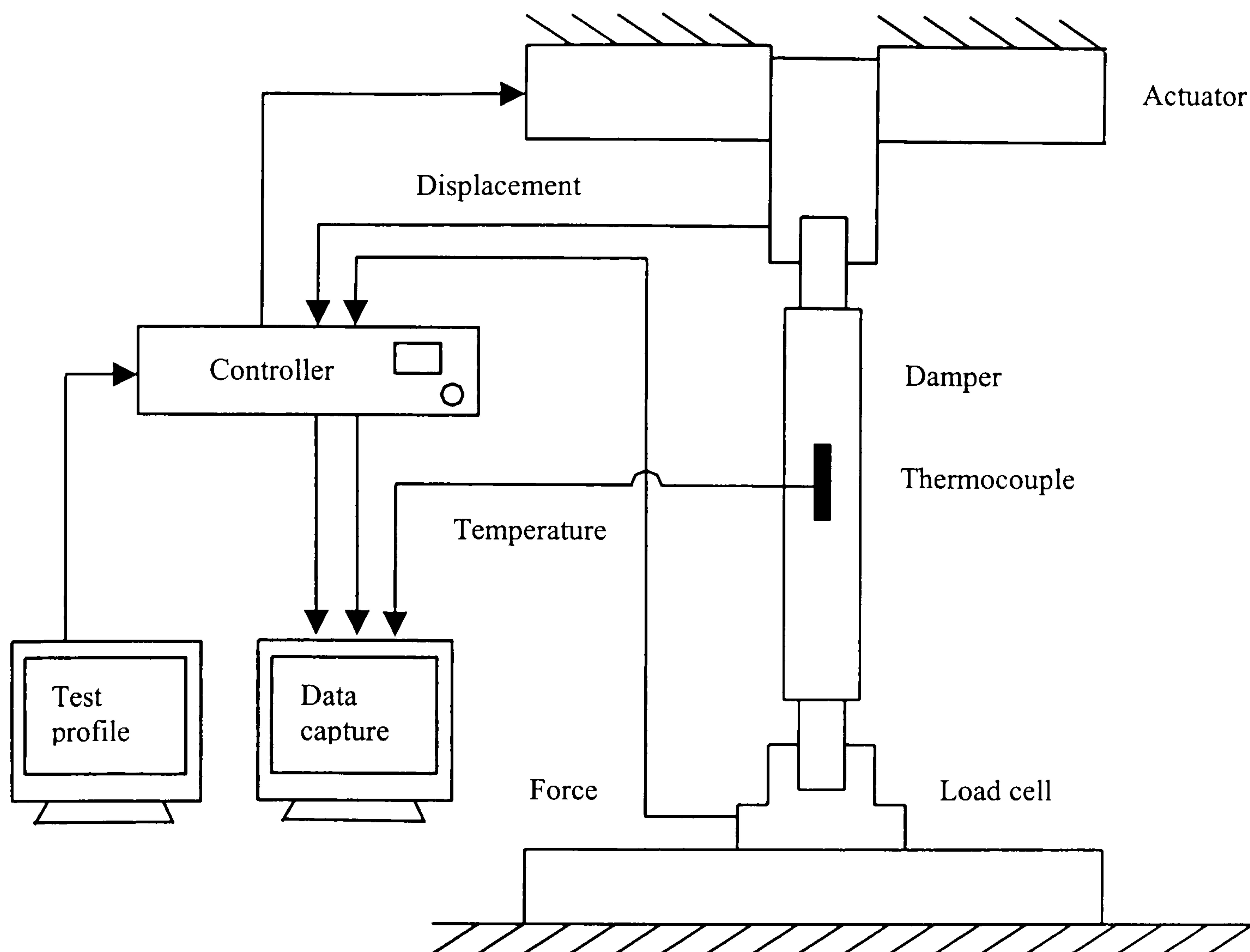


Fig. 3.2.2 Diagram of Damper Test Rig and instrumentation.

The piston of the damper was connected to the actuator head and the main body of the damper was connected via a piezoelectric force transducer to the surrounding metal structure. A thermocouple was attached to the body of the damper to monitor its temperature. Due to temperature dependency of the damping characteristic, care had to be taken. For example when the damper was subjected to sinusoidal inputs at various frequencies, only a limited number cycles were performed for each frequency and the temperature of the damper body was kept within 20 – 21 °C during the test. Three signals were captured using another PC. The captured signals were temperature, damper force and displacement, the displacement was measured by the LVDT embedded within the actuator. All signals were pre-filtered with cut-off frequency of

30 Hz and then sampled at 500 Hz, a high enough rate to ensure that the integration and differentiation routines should produce accurate result [5]. Once the data had been obtained, the static-offset was removed. The velocity and acceleration signals can then be determined by numerical differentiation of the measured displacement signal.

3.2.2 Experimental Data

The experimental data have been gathered, and are presented in the forms of 'work diagrams' and 'characteristic diagrams'. A work diagram is a plot of the damper force as a function of the displacement, and a characteristic diagram is a plot of the damper force as a function of the velocity. As examples, work and characteristic diagrams obtained with sine wave input excitation at different amplitudes with the input frequencies set at 1, 5, 10, 15, and 20 Hz are shown in Fig. 3.2.3-Fig.3.2.7. The work diagram is in part (a) of the figures, while the characteristic diagram is in part (b). For each plot, small, medium and high amplitudes are shown in dash-dotted-line, solid-line, and dash-line respectively.

First, considering the damper response to sine wave excitation at 1 Hz in Fig. 3.2.3 (a) and (b), it can be seen that the damper characteristic is not symmetrical and hence nonlinear. The discontinuities towards the ends of the strokes, near zero velocity, are in part due to friction.

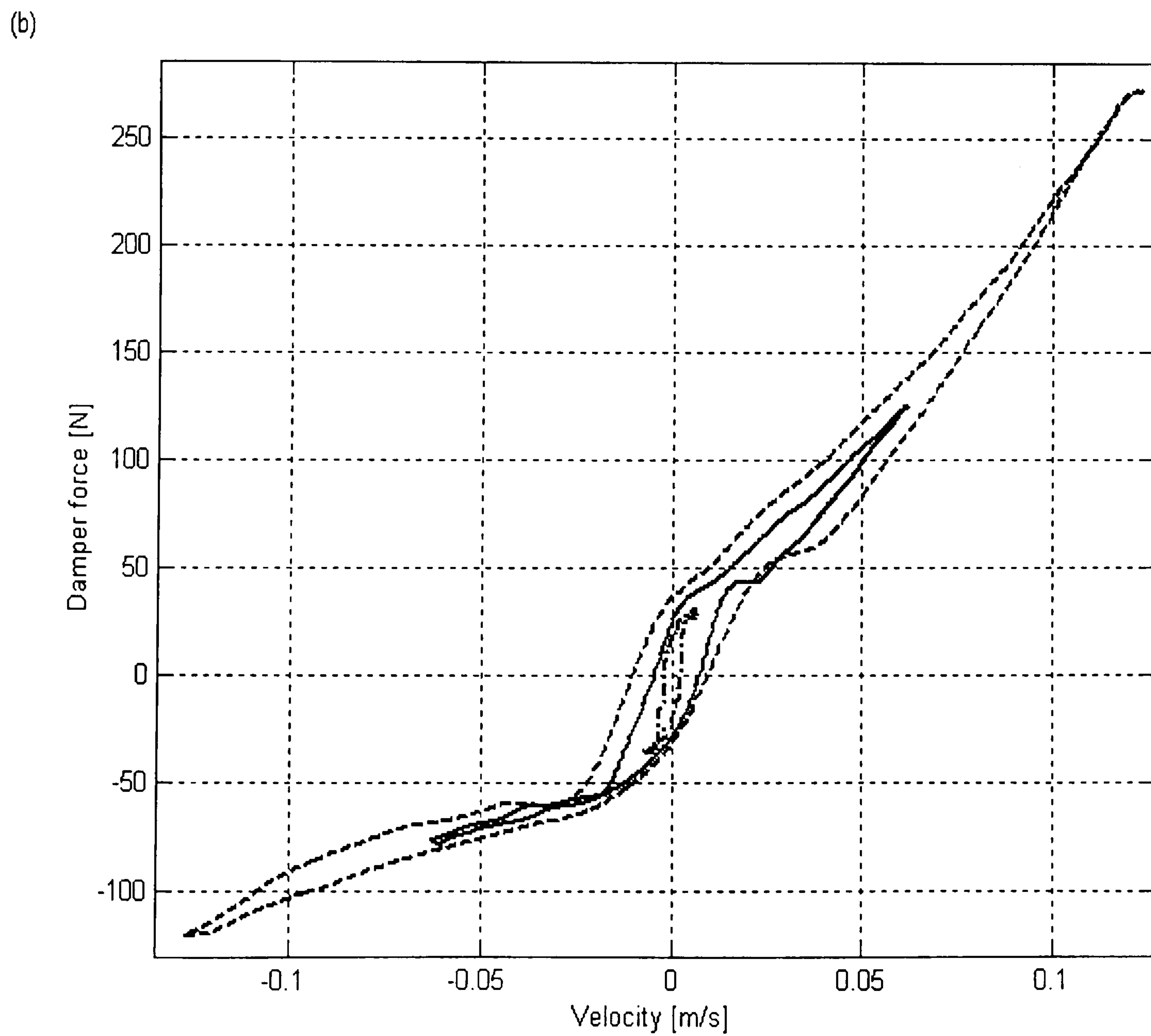
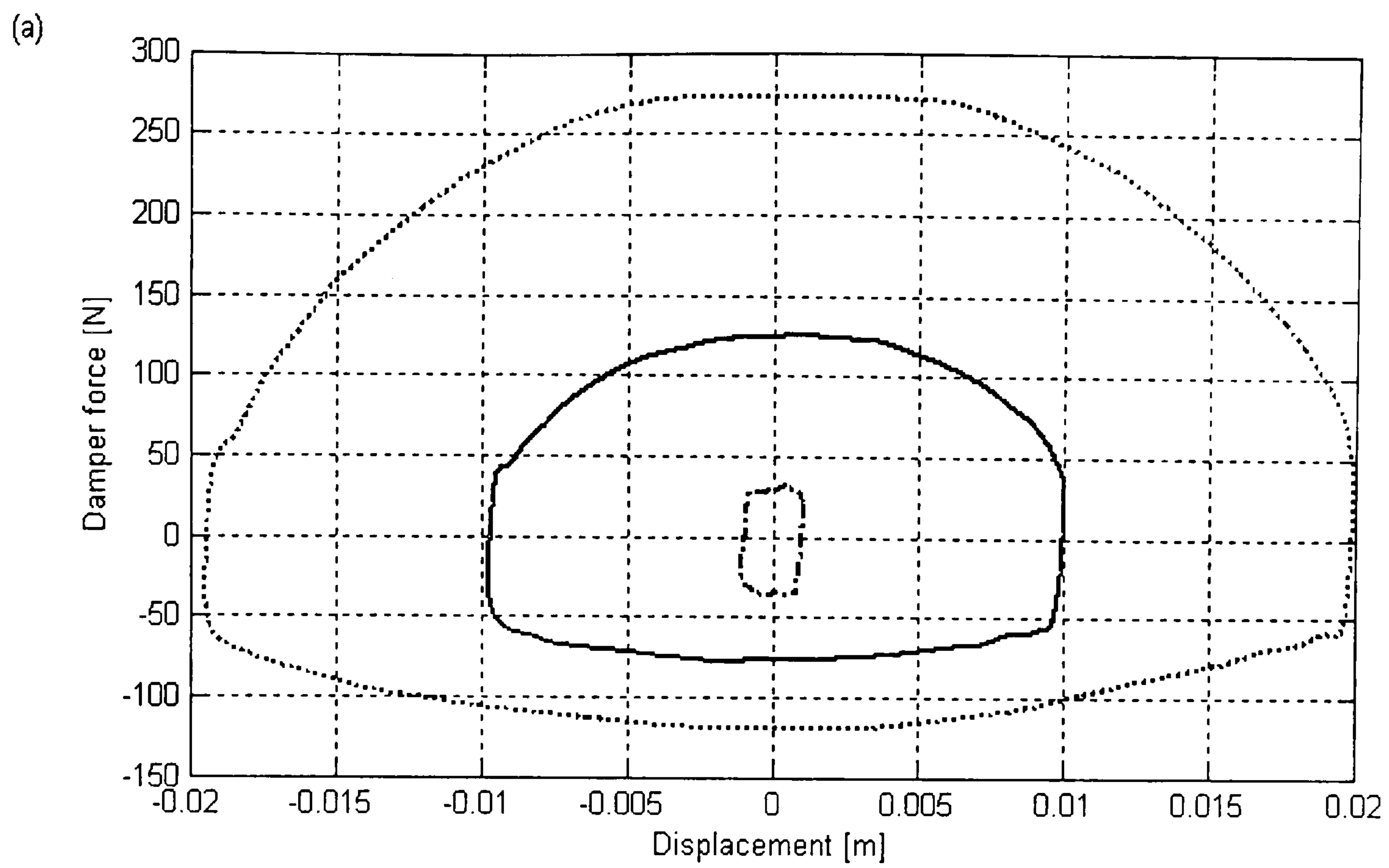


Fig. 3.2.3 Plots of force-displacement and force-velocity, with input frequency of 1 Hz at different amplitudes.

Considering one direction, the linear behaviour, besides the near-zero-velocity discontinuity, is typical for the laminar flow at low velocities of the working fluid. The energy dissipated per unit cycle may be calculated from the area under force-displacement curve. The damping rate or damping coefficient can be obtained as the average ratio of force over velocity. From the two plots, energy dissipated in the rebound (positive damper force, see Fig. 3.2.3 (a)) is approximately 2-2.5 times of that of in the compression direction. The work diagram would have been an ellipse, and the characteristic diagram would have been a straight line if the damper had the behaviour of a linear viscous damper and no friction. For one particular velocity, different values of damper force were obtained, depending upon the piston location or direction of the stroke. This caused a loop, sometimes improperly called hysteresis, to be seen, see Fig. 3.2.3 (b). The loop becomes bigger as the amplitude of excitation gets larger. As the excitation amplitude increases, a larger amount of friction, about ± 50 N, is seen, and the damping force in the rebound direction becomes larger than that of the compression direction.

As the frequency increases, see for example Fig. 3.2.4 and Fig. 3.2.5, the loop, hence the hysteresis become larger. The friction is evident at low amplitude, while it is progressively hidden as the amplitude increases. The overall shapes of both work and characteristic diagrams change when compared with those of Fig. 3.2.3. The asymmetrical characteristic and friction become less and less important at high frequencies, see Fig. 3.2.6 and Fig. 3.2.7. The damper characteristic appears to be linear with a bigger loop size, hence a greater amount of hysteresis effect, as well as 'offset' from the point of zero force and velocity, see Fig. 3.2.7 (a) and (b).

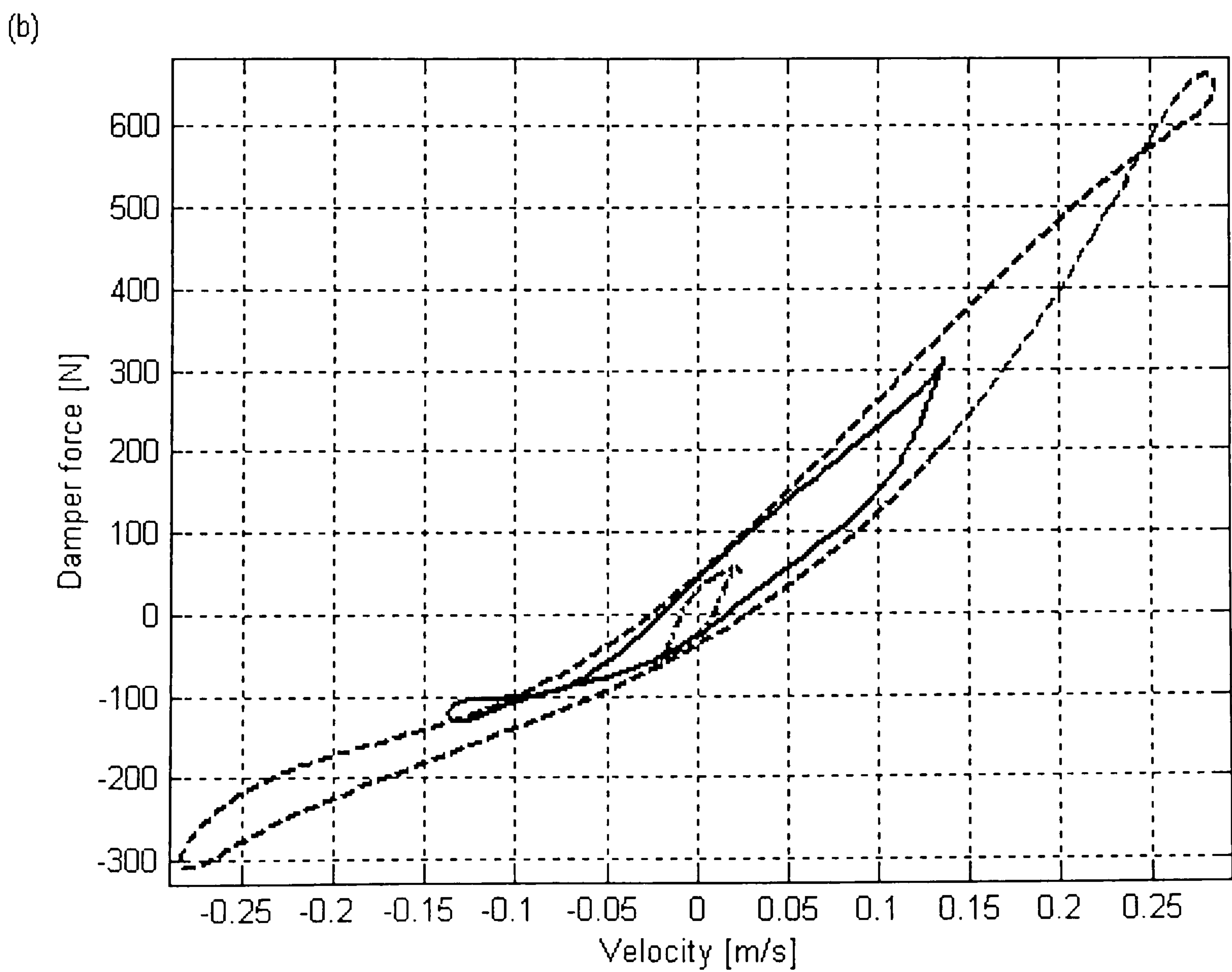
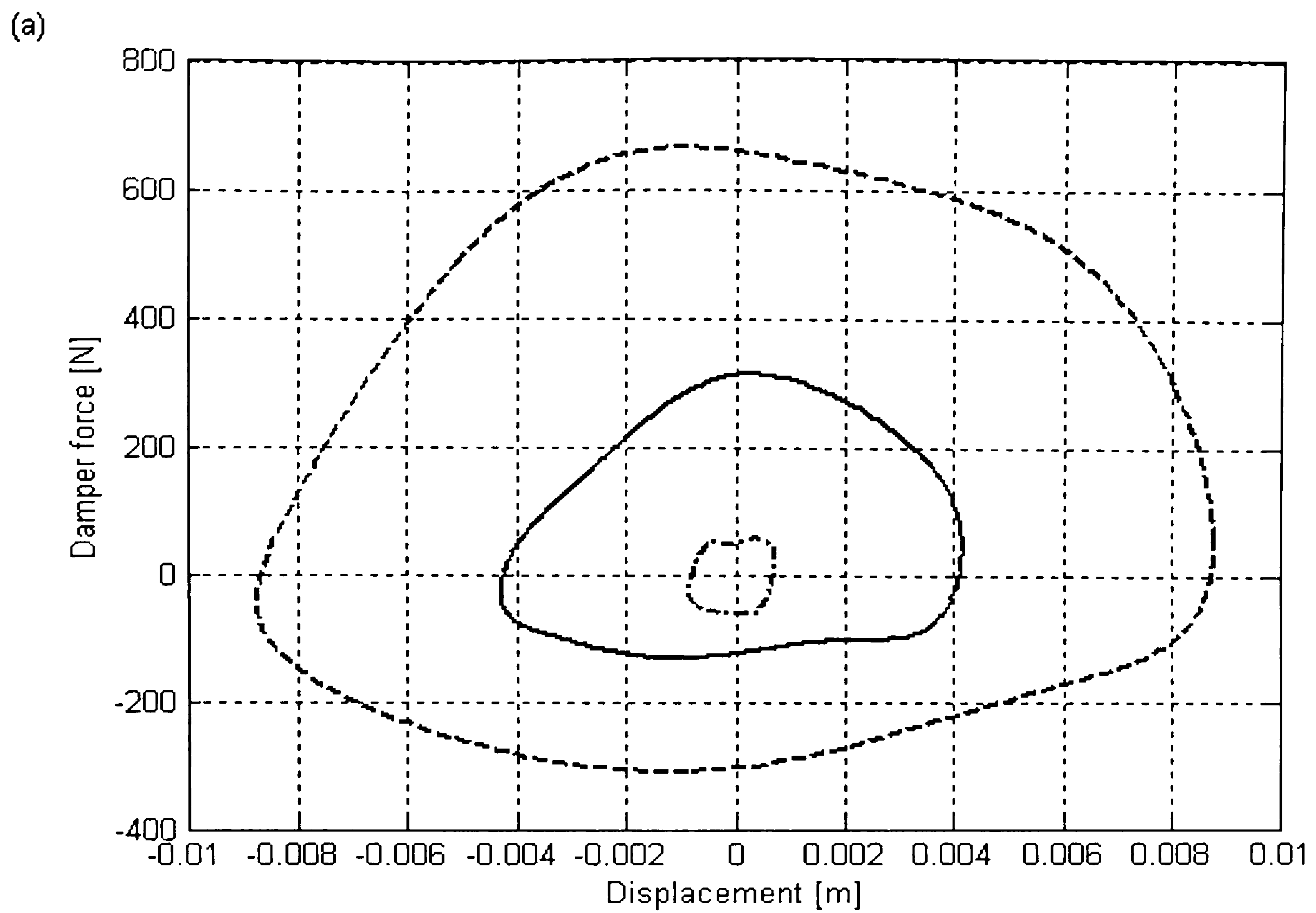


Fig. 3.2.4 Plots of force-displacement and force-velocity, with input frequency of 5 Hz at different amplitudes.

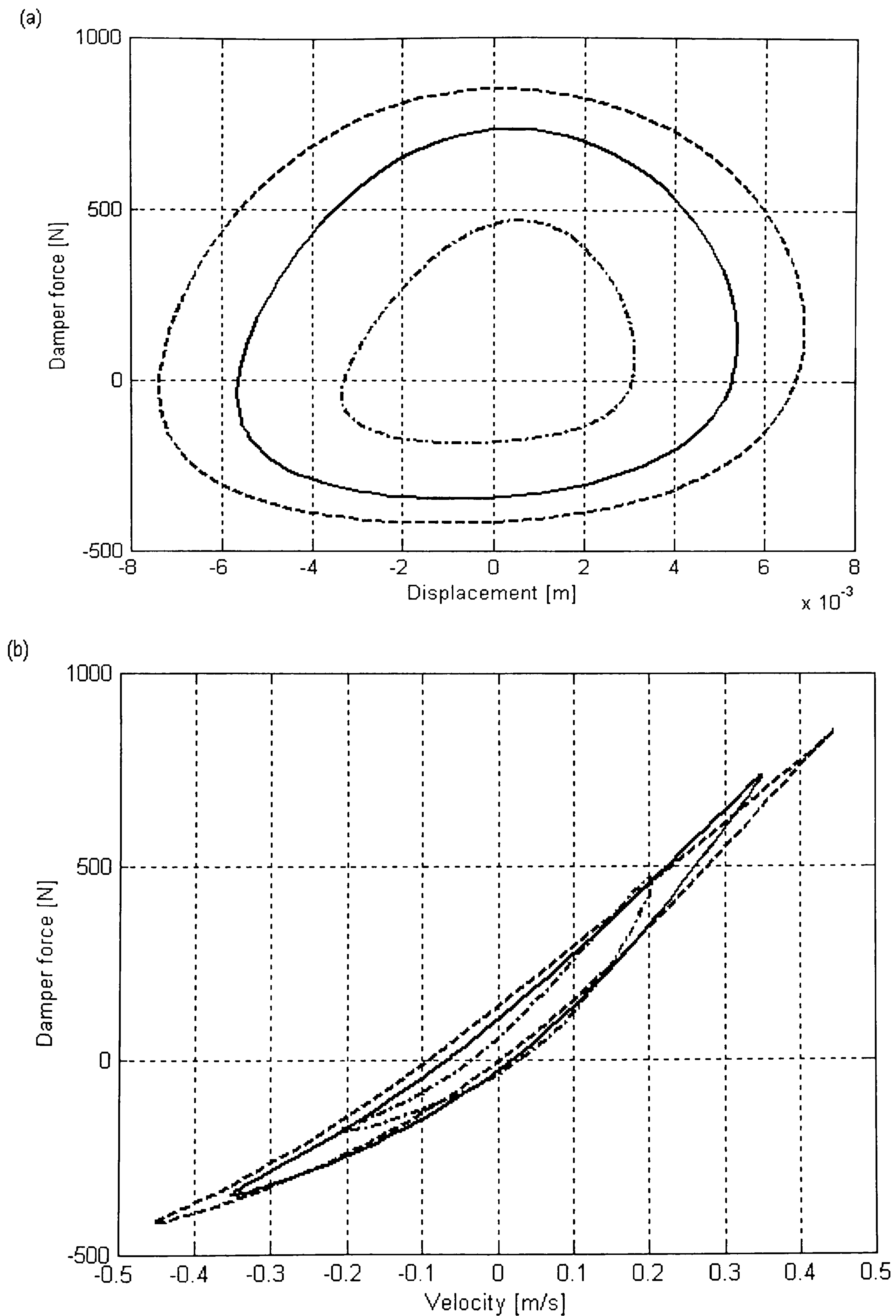


Fig. 3.2.5 Plots of force-displacement and force-velocity, with input frequency of 10 Hz at different amplitudes.

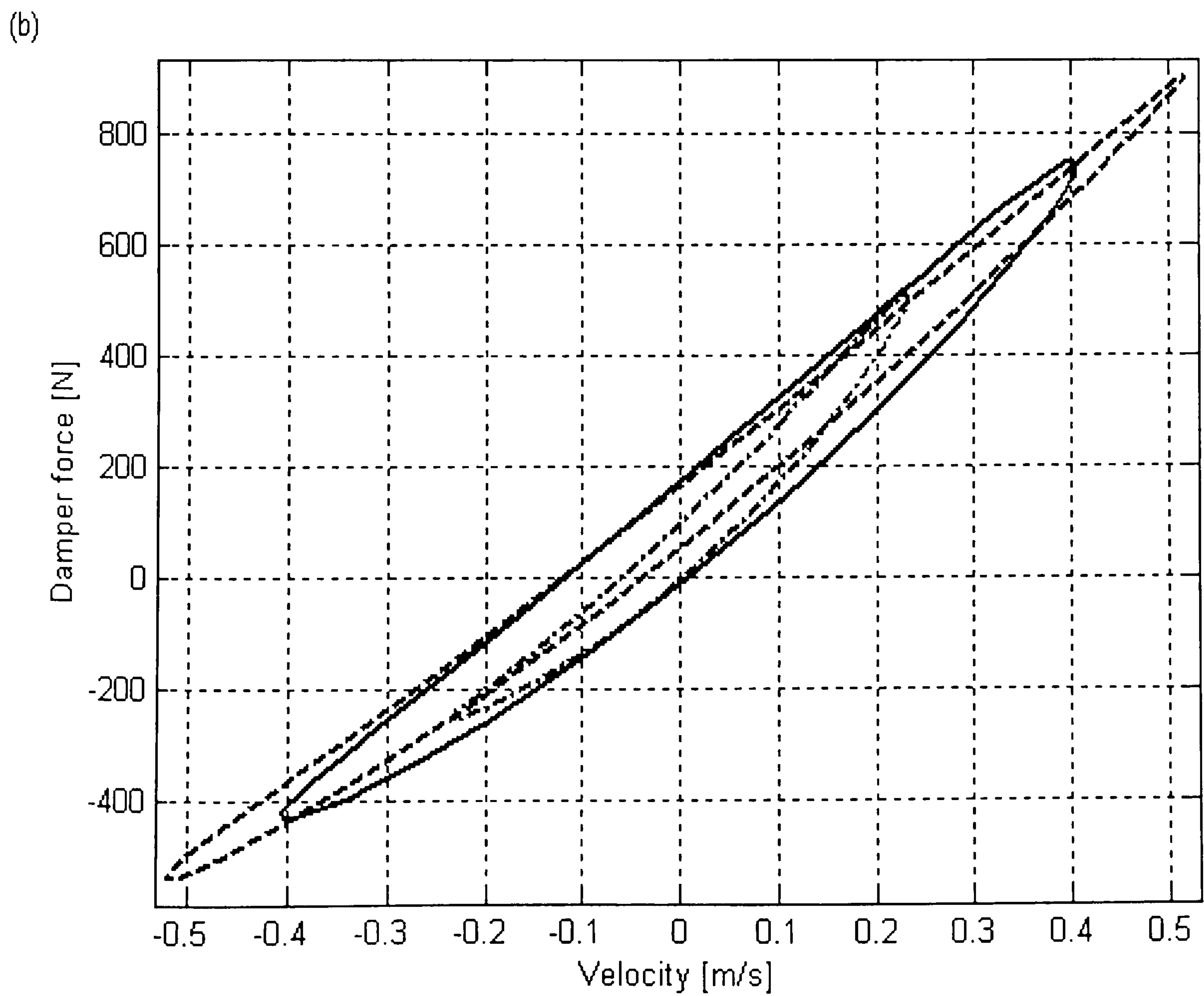
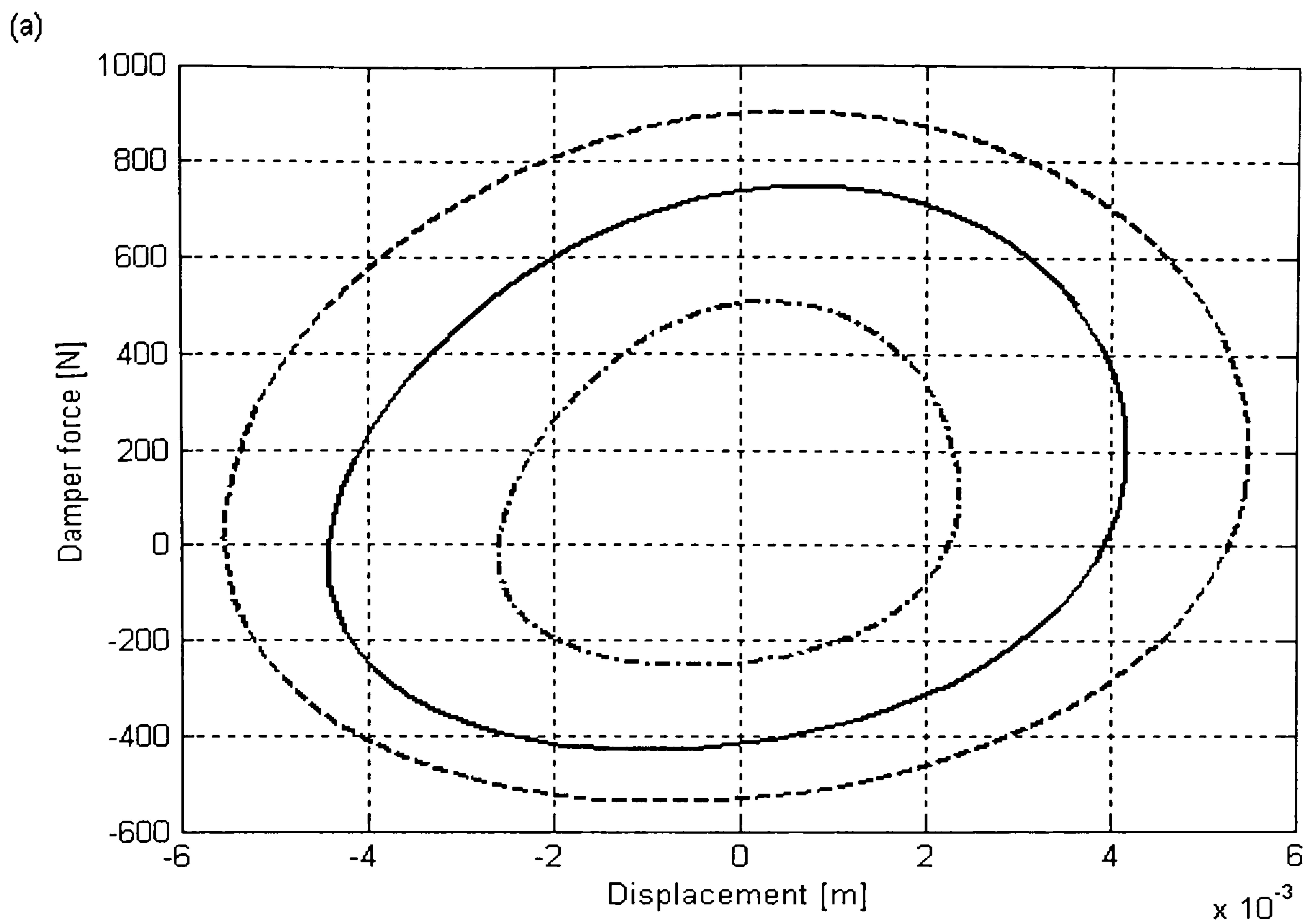


Fig. 3.2.6 Plots of force-displacement and force-velocity, with input frequency of 15 Hz at different amplitudes.

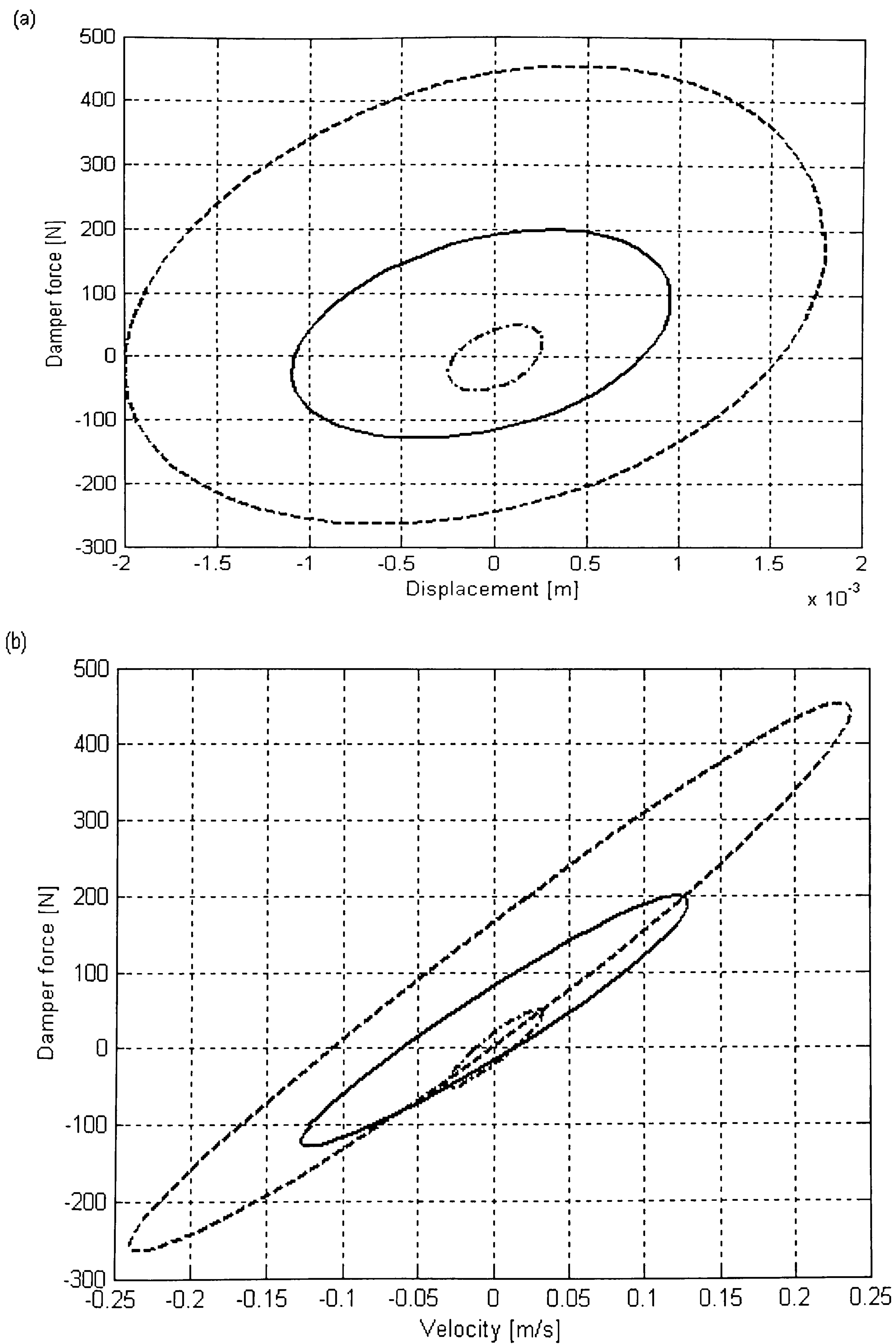


Fig. 3.2.7 Plots of force-displacement and force-velocity, with input frequency of 20 Hz at different amplitudes.

This phenomenon suggests that the force is not perfectly in phase with velocity and that the damper does not behave only as a damping device, but also stores elastic energy.

A test result obtained from a swept sine test is presented in Fig. 3.2.8. The excitation signal was increasing in frequency from 0.5 Hz up to 30 Hz, and decreasing in amplitude. Three intermediate sections are shown here for ease of observing; low, medium and high frequencies. These give an overall view of the damper's behaviour when excitation frequency is progressively increased. Two important points deserved to be emphasised these are, first, at low frequencies the asymmetrical behaviour and the friction are important and are the main features observed. When the excitation frequency increases, the friction become less important and the amount of hysteresis become more dominant. At intermediate frequencies, friction is still seen but has becoming less obvious, while the size of the loop becomes larger. At higher frequencies, the damper behaviour is shown in Fig. 3.2.8(c). At the chosen frequency in Fig. 3.2.8(c), the excitation amplitude and velocity are relatively small. Apart from the loop and offset, the measured force appears to be a linear function of the displacement and velocity across the damper. This suggests that the damper was operating in its linear region at high frequency, and the hysteresis was dominant and the friction was not significant, hence the loops were seen on both the work and characteristic diagrams. Secondly, therefore, the damper behaviour changes from a damping device to an elastic energy-storing device as frequency of excitation increases. These plots of the damper behaviour suggest directions when modelling the damper, within the frequency range of interest.

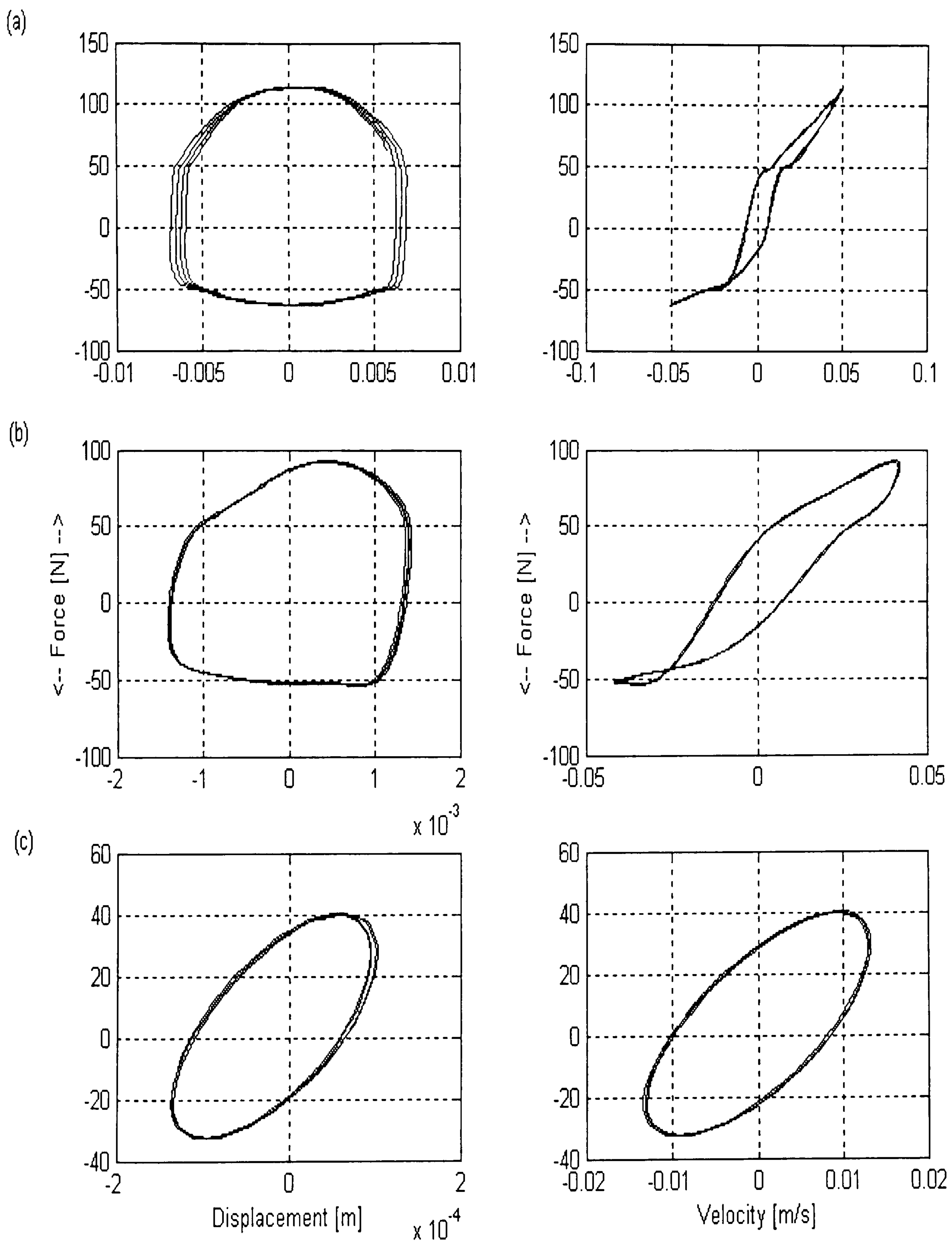


Fig. 3.2.8 Sectioned plots of force-displacement and force-velocity, with swept sine input frequency range of 0.5-30 Hz; (a) low frequencies, (b) intermediate frequencies, (c) high frequencies.

3.3 Damper Models

Real damping mechanisms are usually quite complex. The experimental data presented in the last sub-section has shown that it certainly is the case for automotive dampers. However, in general, engineers may use one or combinations of the following three types of damping elements when modelling vibrating system.

- (1) Viscous damping. A linear dashpot is an idealized viscous damping element where relationship between damping force and velocity across the element is a linear one. And, the force is a function of the velocity variable only.
- (2) Material, solid or Hysteresis damping. A simple model for hysteresis damping is a model in which a spring connected in parallel to a dashpot, so that damping force is a function of both displacement and velocity, as well as frequency of the input.
- (3) Coulomb or dry friction damping. A simple model for Coulomb damping is where damping force is constant and in the opposite direction to the velocity across the element.

An accurate modelling of the dynamic behaviour of a damper using a combination of these simplified elements may give good results for one excitation frequency. However, the modelling task may encounter great difficulty when the dynamic behaviour of interest is not at one particular frequency, but a range of frequencies. The aim here is to develop a model for the damper that gives a good prediction to the measured damper force with the fewest possible parameters in the interest frequency range of 0.5-30 Hz. Starting from a simple model, parameter estimation procedure is carried out to estimate the model parameters and to assess the quality of fit for the model to the measured data. In this section, a review on the modelling of the

automotive dampers is first discussed. The damper models for this case study is then proposed. The performances of the models together with the results of the parameter estimation will then be presented and discussed in section 3.4.

As evident from the experimental data, the dynamic behaviour of an automotive damper is far from simple and thus modelling its characteristics accurately is not an easy task. In fact, the damper is one of the most complex parts of the suspension system to model. Three different approaches may be adopted in modelling the behaviour of a damper. The first approach is when the damper's behaviour is modelled based on its detailed physical construction. Various mechanisms, which contribute to the damper force, are identified and modelled using laws of fluid flow and thermodynamics. This approach can be found in the literature, for example [2,103-106]. In [103,104], the model, which was a highly nonlinear one with 82 parameters, and showed good agreement between the analogue computer simulated data and the experimental data. Two main disadvantages of this model are that, models of this type are by their very nature specific to each damper, and because of its complexity, the model cannot be readily identified from the damper force measured. The internal pressures must be measured or predicted numerically and this involves many time consuming procedures. They are therefore not suited for use in full vehicle simulation. Hall et al [107] used this model in a quarter car vehicle simulation and established that typical linear and bilinear models of damper are inaccurate and lead to over optimistic estimates of the ride performance. Reybrouck [105] proposed a readily identifiable model, which gave an explicit expression for the damper force as a function of displacement, velocity and acceleration, and contained 14 parameters. The parameters of the model were identified by minimizing a cost function formulated

from the squared difference between the measured and modelled damper force. Minimising is performed using the Levenburg-Marquardt algorithm implemented in MATLAB [108]. However, as far as the hysteresis is concerned, the model is defective [101]. Duym et al [101] presented two physical models that are able to extract the internal valve parameters from data without hysteresis. Wallaschek [106] adopted the application of harmonic and stochastic linearisation procedures to the damper dynamics. The model was equivalent to a piston moving in a closed cylinder containing a fluid, which could pass between the rebound and compression chambers through an orifice of constant cross section in the piston head. The energy dissipation due to viscous flow and throttle losses, as well friction and backlash, were included in the model. The final model required nine coefficients that were only good for a particular frequency, for a different frequency a different sets of coefficients are needed. Surace [2] improved Wallaschek model to include hysteresis, however the model still required different sets of coefficients for different frequencies.

A second approach is that based on non-parametric models. This is similar to polynomial modelling. The individual terms or parameters do not have physical meaning but the collective contribution of all the terms gives a strong correlation between the modelled force and the measurements. In order to implement a model that copes with hysteresis, most models require the numerical solution to a set of non-linear differential equations. The use of an alternative restoring force method can get round the time consuming activity of the iterative simulation and identification routines, see for example [109-113]. One limitation of this approach is that the model is valid only within the boundary of the test condition. This means, a model that has been developed using smooth road test data may not be accurate for use under rough

road condition. It has been shown [109-112] that the force-state maps in the restoring force model was frequency dependent. This means, different maps were required for different frequencies. When taking into account the effect of temperature, a larger collection of force-state maps was needed [113].

An alternative to modelling a damper by physical details is to use a simplified model with elements such as a dashpot, spring, backlash and friction elements in various combinations. A model, which included hysteresis and backlash as well as friction and asymmetric characteristics, was presented by Karadayi and Masada [114]. The model had six parameters, and a good fit was obtained at an amplitude of 55 mm and frequency of 0.2 Hz, but no results were presented for higher frequencies. Besinger et al [115] modelled a damper for heavy commercial vehicle using a tri-linear dashpot connected in series with a cubic spring. This model also contained six parameters. Besinger et al used a Rung-Kutta algorithm for simulation and the identification was performed using trial and error because a convergence difficulty was encountered due to local minima.

It was suggested by the experimental data and the discussion made in Section 3.2.2 that over the frequency range of interest, 0.5-30 Hz, the damper behaviour progressively changed from a viscous damping device to a energy storing device. Secondly, the contribution of friction and hysteresis, as excitation frequency increased from 0.5 Hz to 30 Hz, change in a complex manner. In this investigation, the damper was therefore modelled from simplified elements such as a dashpot and a spring. The model however did not include friction and hysteresis. A separate investigation on the

estimation of a more complex damper model, which includes friction and other nonlinear behaviour, is provided in Appendix A.

For the investigation here, to select a model that gives a good prediction to the measured damper force with the fewest parameters, an iterative procedure of selecting the model, estimating the model parameters and assessing the quality of fit as described in general system identification/parameter estimation case in Chapter 1 was carried out. Table 3.3.1 provides the list of models considered. They are derived from two simplified configurations shown in Fig. 3.3.1 (a) and (b).

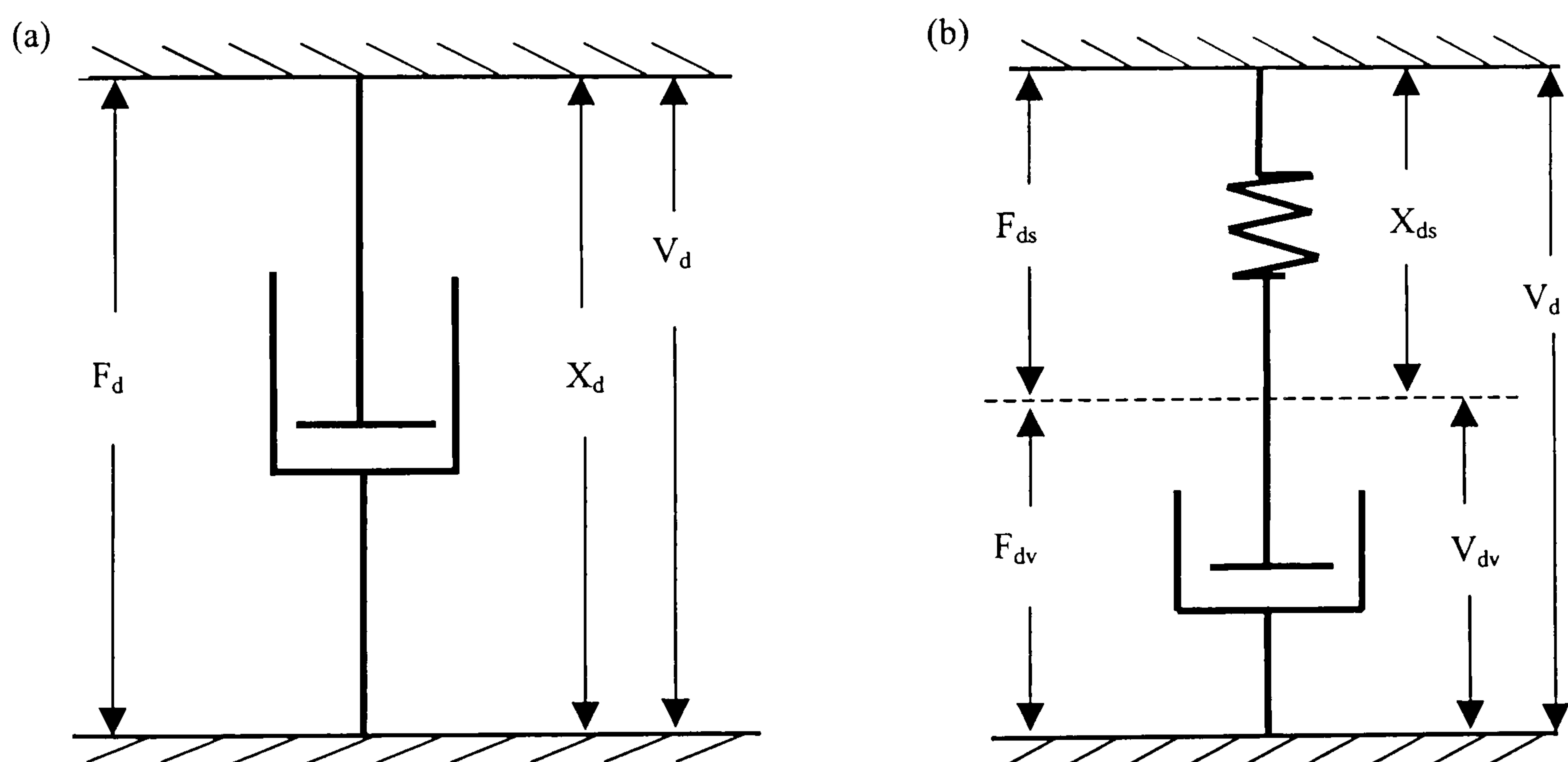


Fig. 3.3.1 Simplified damper models; (a) a dashpot, (b) a spring in series with a dashpot.

Model 1 and Model 2 are linear and bilinear models where the damper force, f_d , is a function of the velocity across the damper, v_d , only, see equation (3.1).

Model 1	Linear dashpot	$\mathbf{f}_d = d_d \mathbf{v}_d$	(3.1)
Model 2	Bilinear dashpot	$\mathbf{f}_d = \left(d_r \left(\frac{1 + \text{sign}(\mathbf{v}_d)}{2} \right) + d_c \left(\frac{1 - \text{sign}(\mathbf{v}_d)}{2} \right) \right) \mathbf{v}_d$	(3.2)
Model 3	Linear spring in series with bilinear dashpot	$\mathbf{f}_d = \mathbf{f}_{ds} = \mathbf{f}_{dv}$	(3.3)
		$\mathbf{f}_{ds} = k_{c1} \mathbf{x}_s$	(3.4)
		$\mathbf{f}_{dv} = \left(d_r \left(\frac{1 + \text{sign}(\mathbf{v}_d)}{2} \right) + d_c \left(\frac{1 - \text{sign}(\mathbf{v}_d)}{2} \right) \right) \mathbf{v}_d$	(3.5)
Model 4	Cubic spring in series with bilinear dashpot	$\mathbf{f}_{ds} = k_{c1} \mathbf{x}_s + k_{c3} \mathbf{x}_s^3$	(3.6)
		$\mathbf{f}_{dv} = \left(d_r \left(\frac{1 + \text{sign}(\mathbf{v}_d)}{2} \right) + d_c \left(\frac{1 - \text{sign}(\mathbf{v}_d)}{2} \right) \right) \mathbf{v}_d$	
Model 5	Cubic spring in series with bilinear dashpot (with transition factor)	$\mathbf{f}_{ds} = k_{c1} \mathbf{x}_s + k_{c3} \mathbf{x}_s^3$	(3.7)
		$\mathbf{f}_{dv} = d(\mathbf{v}_d) \mathbf{v}_d$	
		$d(\mathbf{v}_d) = \frac{r_1(\mathbf{v}_d - \alpha)}{\alpha \sqrt{1 + \left(\frac{\mathbf{v}_d - \alpha}{\alpha} \right)^2}} + r_2$	
		$r_1 = (d_r - d_c)/2, \quad r_2 = (d_r + d_c)/2$	

Table 3.3.1 List of damper models considered.

The equivalent damping rate of the damper is denoted by d_d for the linear model, while d_r and d_c denote damping rates in rebound and compression directions respectively. The displacement and velocity cross the damper are denoted by x_d and v_d respectively, see Fig. 3.3.1 (a). The model with equivalent linear or bilinear damping and stiffness is useful when the model is to be used in fast simulation study or control purposes, where accurate prediction of damper behaviour is not a prime concern. For example, with a spring element included, Model 1 may be used to obtain the equivalent stiffness and damping of an automotive damper when excited by a simulated road test profile. The damper model may then be used as part of a study on vehicle behaviour in cornering and/or accelerating and braking motions. A combination of a spring element in parallel with a dashpot is not included in the presentation here, as its fitting result were poor.

Model 3 comprises of a linear spring in series with a bilinear dashpot, as depicted in Fig. 3.3.1 (b). The viscous damping force produced by the dashpot element is denoted by f_{dv} , see equation (3.4) –(3.5). The compliance spring force, f_{ds} , is the product of the spring stiffness, k_{cl} , and the displacement across the element, x_{ds} . The model is justified from the fact that when the damper is excited by a force-input, the initial response is dependent on the stiffness of the mixed gas-oil solution. Afterwards the damper rod will move with constant velocity. This behaviour is typical for a series combination of a spring and a dashpot element. The damper force is then a function of both displacement and velocity. Inclusion of a spring element in Model 3 attempts to account for the compliance of the damper, which tends to be significant when the excitation frequency increases. In stead of a linear compliance, Model 4 employs a cubic spring using equation (3.6) to further account for non-linear compliance in the

damper. The model then contains four parameter to be identified; d_r , d_c , k_{c1} and k_{c3} . A further modification from Model 4 is to reduce a sudden change of damping rate between that in rebound and in compression. A transition factor, α , is introduced in Model 5, see equation (3.7) and (3.8). The damping rate is then a function of velocity signal. Model 5 is similar to that in [115] but only has five parameters; d_r , d_c , k_{c1} , k_{c3} , and α .

In the next section, the quality of each model on fitting the measured data will be presented and discussed. The performance of the four numerical search methods for estimating each model parameters is then compared.

3.4 Parameter Estimation

3.4.1 Estimation method

Having chosen a set of candidate models for the damper, the next important step is to estimate the parameters of each model so that the quality of fit of the model to the experimental data can be assessed. The estimation was formulated as a single-objective optimisation problem where the difference between the measured and modelled damper forces is to be minimised. The minimisation was a nonlinear programming problem in which iterative search for best model parameters were required. Fig. 3.4.1 depicts the general concept of the estimation technique used.

Having collected a required batch of data from the damper testing and after some data post-processing, the parameter search procedure was carried out as in the dotted box shown in Fig. 3.4.1. The dotted box separates and indicates that the estimation technique used was an ‘off-line’ processing. The signals such as the displacement and velocity are fed to the damper model to generate the model outputs. The modelled damper force together with the measured force was then used to form the cost function in the minimisation process.

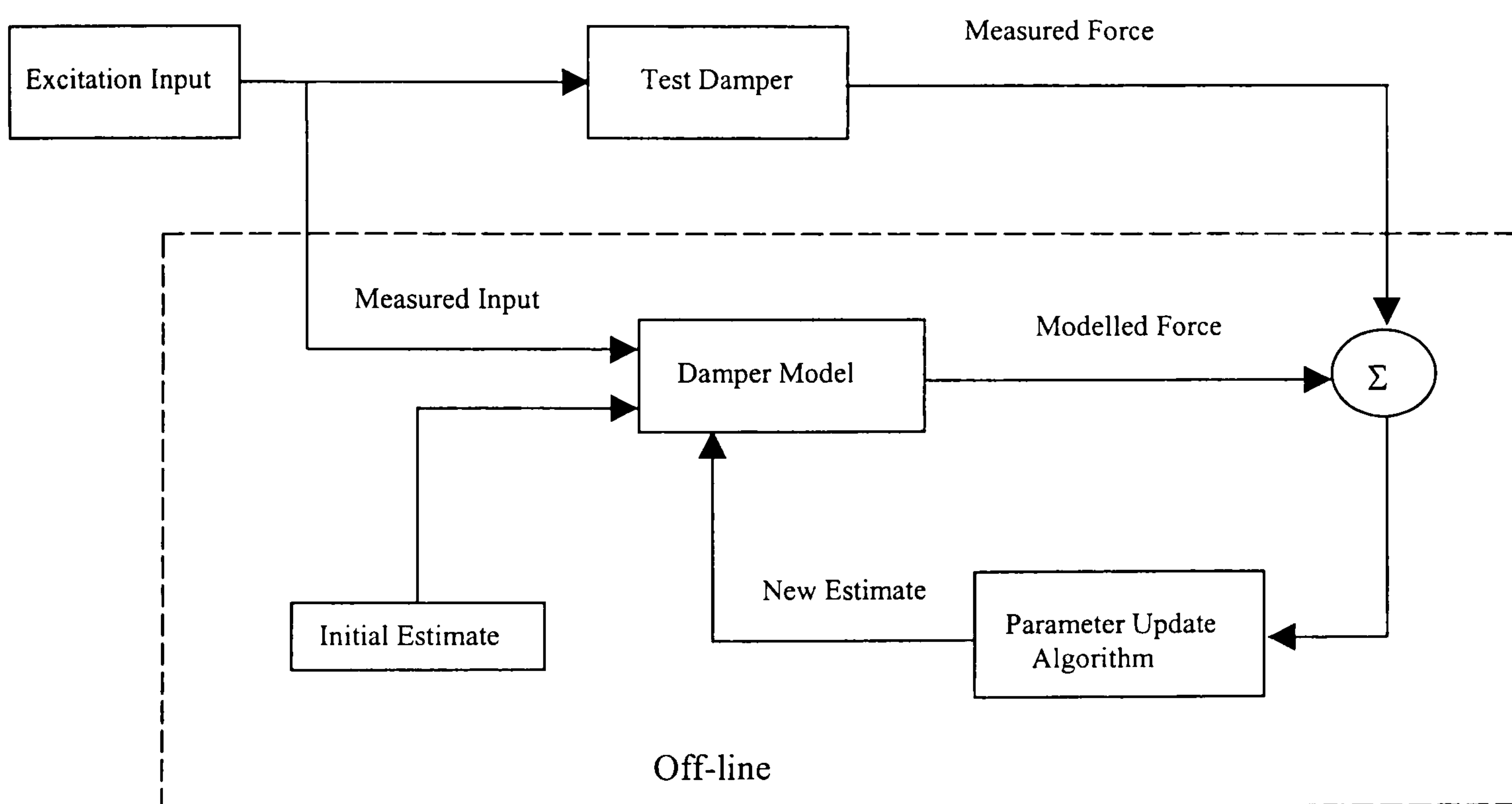


Fig. 3.4.1 Parameter estimation technique.

The cost function for a search algorithm to work on when searching for the best possible parameter values was chosen to be the normalised mean squared error (MSE) of the modelled and measured forces. The choice is a standard one, which is convenient both for computation and analysis. The MSE is commonly used as a measured of how well the model outputs agree with the system outputs [116] ; a MSE

of less than 5.0 percent indicates good agreement while one of less than 1.0 reflects an excellent fit. The MSE value is defined by

$$\text{MSE}(\mathbf{f}_d) = \frac{100}{N \sigma_{f_d}^2} \sum_{i=1}^N (\mathbf{f}_{d_i} - \hat{\mathbf{f}}_{d_i})^2 \quad (3.9)$$

where $\sigma_{f_d}^2$ is the variance of the measured damper force \mathbf{f}_d , $\hat{\mathbf{f}}_d$ is the modelled damper force and N is number of data samples. The search algorithm starts each search run with one or a population of initial parameters. The algorithm then updates the estimated parameters and moves towards a better region where the optimum point is thought to be located. The optimal parameter is then an individual that optimises (minimises) the MSE value. All computational codes were implemented in MATLAB [108].

Among the five model considered, the performance of a model is discussed and compared to the others, and the best model is then selected. The ‘best’ model is one which best compromises the conflict between model quality of fit and model complexity including the issues such as modelling effort required and the time consumed in the estimation process. For each of the five models, the model parameters are identified by four different numerical search methods, the GB, DS, DE and dvHDE. As the complexity of the parameter estimation problem increases from Model 1 to Model 2, the performance of each numerical search method is analysed and compared to the others. The results for these are presented in the next section.

3.4.2 Estimation Results

3.4.2.1 Model quality

Though offering simplicity and fast identification, Model 1 poorly predicted the measured damper force. The quality of fit in term of MSE was about 19%. An example of the fitting results is depicted in Fig.3.4.2. The figure presents the plots of work and characteristics diagrams of four instantaneous excitation frequencies. The four sectioned plots represent the damper dynamic behaviour at low-, intermediate- and high-frequency intervals; one plot at low frequency of 0.5 Hz, marked (a), two plots at intermediate frequencies of 1.5 Hz and 4.5 Hz, marked (b) and (c), and one at high frequency of 15 Hz, marked (d). A presentation of the modelled and measured damper force on the same graph for comparison and analysis may give a good overall visual appreciation of how well the model performs. However, presenting the data in sections, i.e. low, medium and high frequency intervals, allows the analysis of the results to be carried out more conveniently.

The linear, equivalent damping rate estimated gives the best linear fit when the measured damper force is modelled as function of velocity only. The result is therefore the best compromise of errors, difference between the modelled and measured forces, from all sampled data points for the whole frequency range of interest. From Fig. 3.4.2, it is obviously seen that, at low and intermediate frequency intervals, the linear dashpot model fails to capture the friction effect at near zero velocities, and the bi-character behaviour elsewhere.

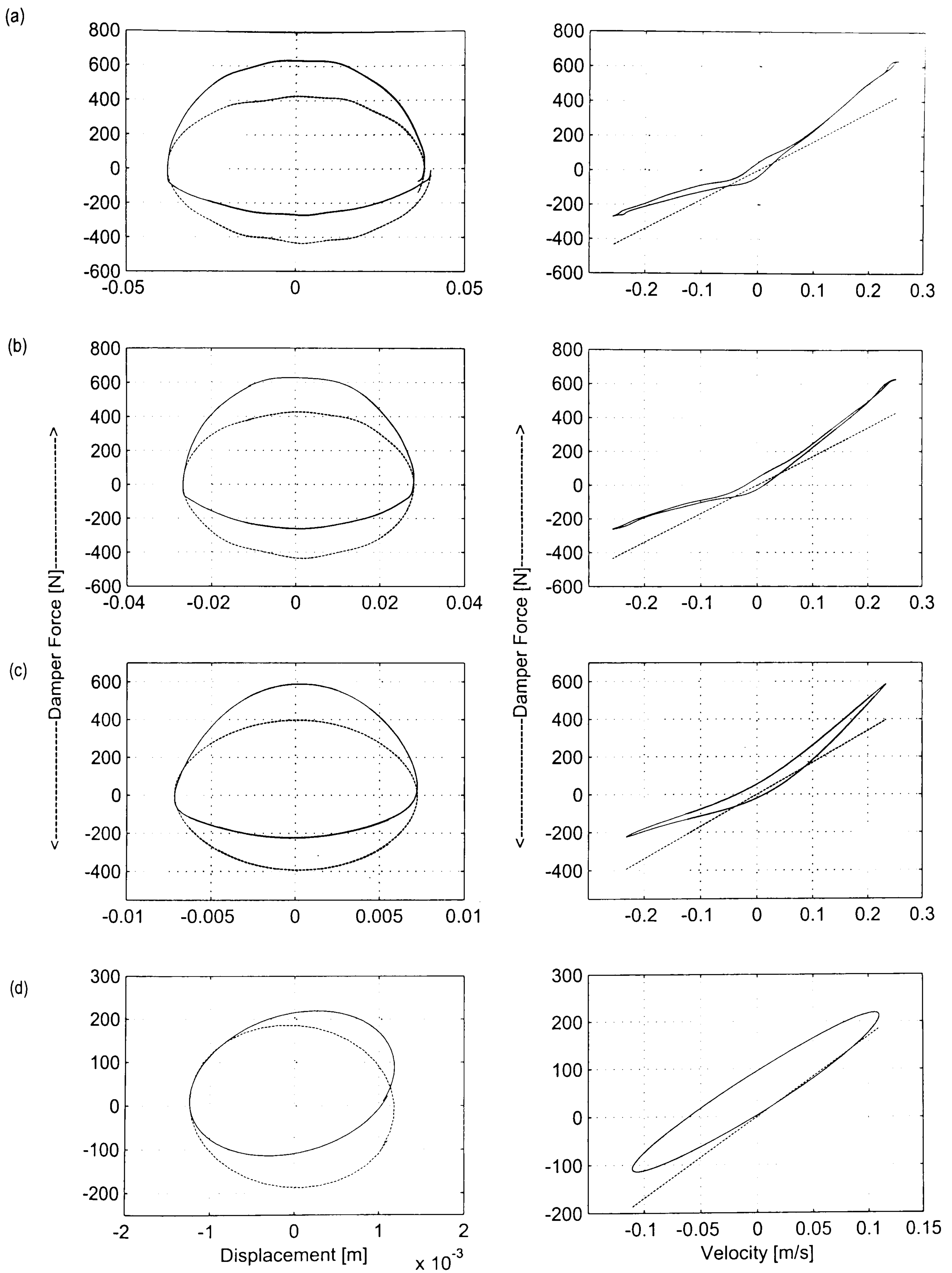


Fig. 3.4.2 Sectioned plots of fitting result obtained from Model 1;
- measured, -- modelled.

At high frequency range, the linear model fit gives an impression, roughly, that the relationship of force and velocity is a linear one. However, the plots also suggests that the force is not a function of velocity only. In addition, other main fitting errors can be seen from the ‘offset’ of damper force and large hysteresis loop, clearly seen on the force-velocity plot in part (d).

The bilinear dashpot model, Model 2, better captures the dynamic behaviour of the test damper, as can be seen from both the work and characteristic diagrams in Fig. 3.4.3. The MSE value for Model 2 is approximately 1.26%. The value has reduced dramatically when compared to that of the linear model. At low and intermediate frequency intervals, the plots of force-displacement show that the model gives a good fit. The plots of force-velocity fit through the ‘hysteresis loop’ in compression and rebound directions, which was not included in the model. The error from nonlinearity and friction can also be seen. While the bilinear character of the model captures relatively well the damper behaviour at low and intermediate frequency intervals, the continuity of ‘bi-linearity’ is unable to match the behaviour at high frequencies, where the damping force appears to be linear with displacement and velocity. In total, the contributions to the fitting errors are nonlinearity and friction, mainly at low frequencies, hysteresis at intermediate and high frequency intervals, which produces size-increasing loop and offset.

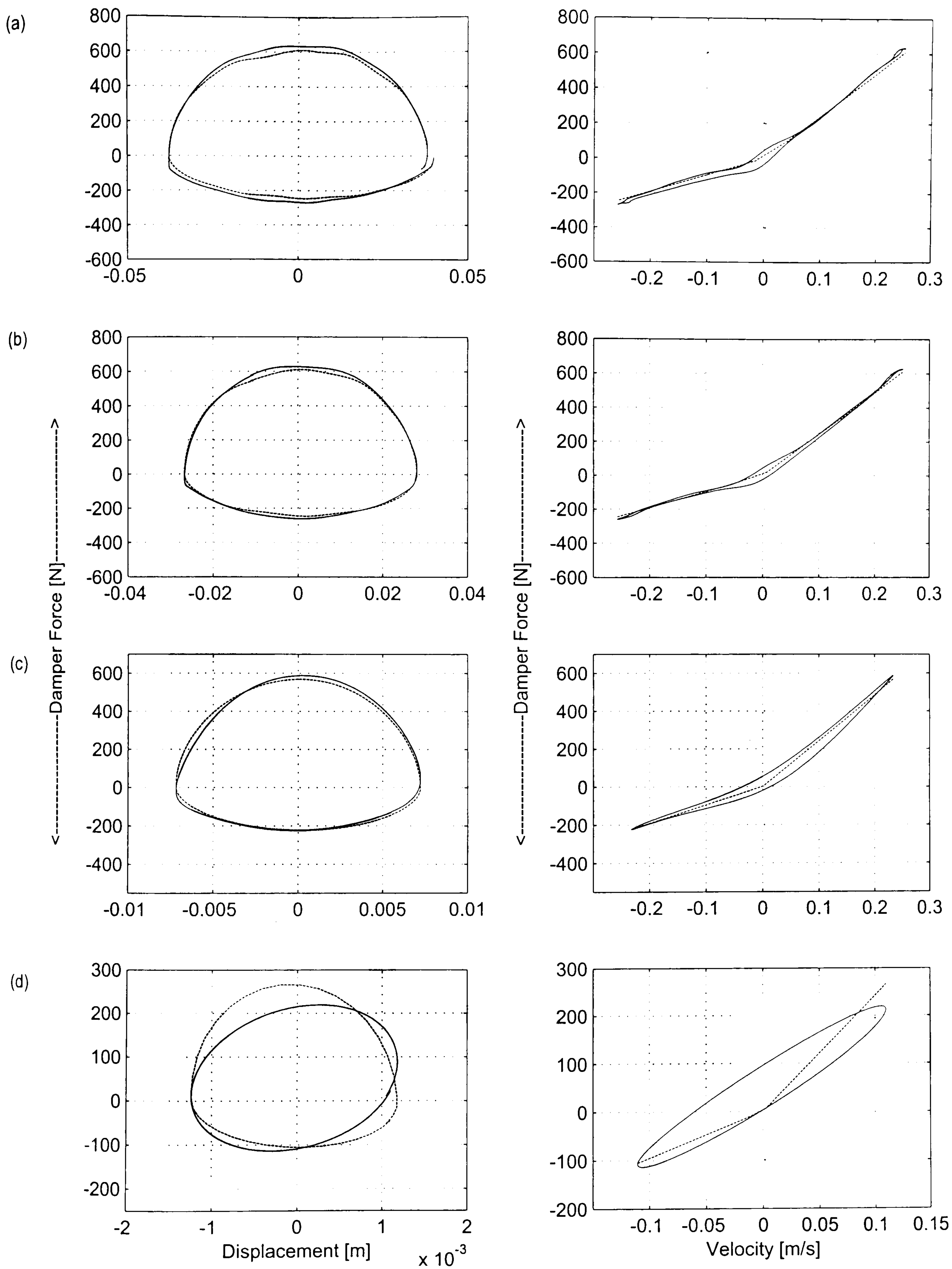


Fig. 3.4.3 Sectioned plots of fitting result obtained from Model 2;
 - measured, -- modelled.

The results from Model 2 show that the damping rate in rebound is approximately 2.5 times of that in the compression direction. The estimated bilinear damping rates are certainly dependent upon where most of the measured data lie. This argument is true for all models considered. There are, for example, more data points in the low velocity intervals than in the high velocity interval for one frequency test. The cost function, which was the normalised sum of the difference between the measured and modelled force, then depends on the distribution of the data points. In this case study, every data point was given the same significant or weight when formulating the measure of fit, i.e. the MSE value. A weighting vector that may be incorporated into the cost function reflects some a-prior knowledge about the data measured, such as measurement noise in the measured input and damper force. However, to obtain the ‘right’ weighting vector is not an easy task. A simple and natural choice is to use a weighting vector of ones, i.e. same weighting for every point.

With Model 3, the test damper was modelled as a linear spring in series with a bilinear dashpot element. The model gives the best fitting result of MSE 0.54%. A three-dimensional comparison plot of the displacement, velocity and damper forces is presented in Fig. 3.4.4, again, to give a better visualisation, only selected sections of data are plotted. It can then be seen from the plot that the model captures the damper’s behaviour well. A detailed selection of plots of force against displacement at low and intermediate frequency intervals shows good quality of fit. The corresponding force-velocity plots show that, besides the friction effect, the model is able to capture the bilinear characteristic and the compliance loop with some accuracy, see Fig. 3.4.5 (a)-(c).

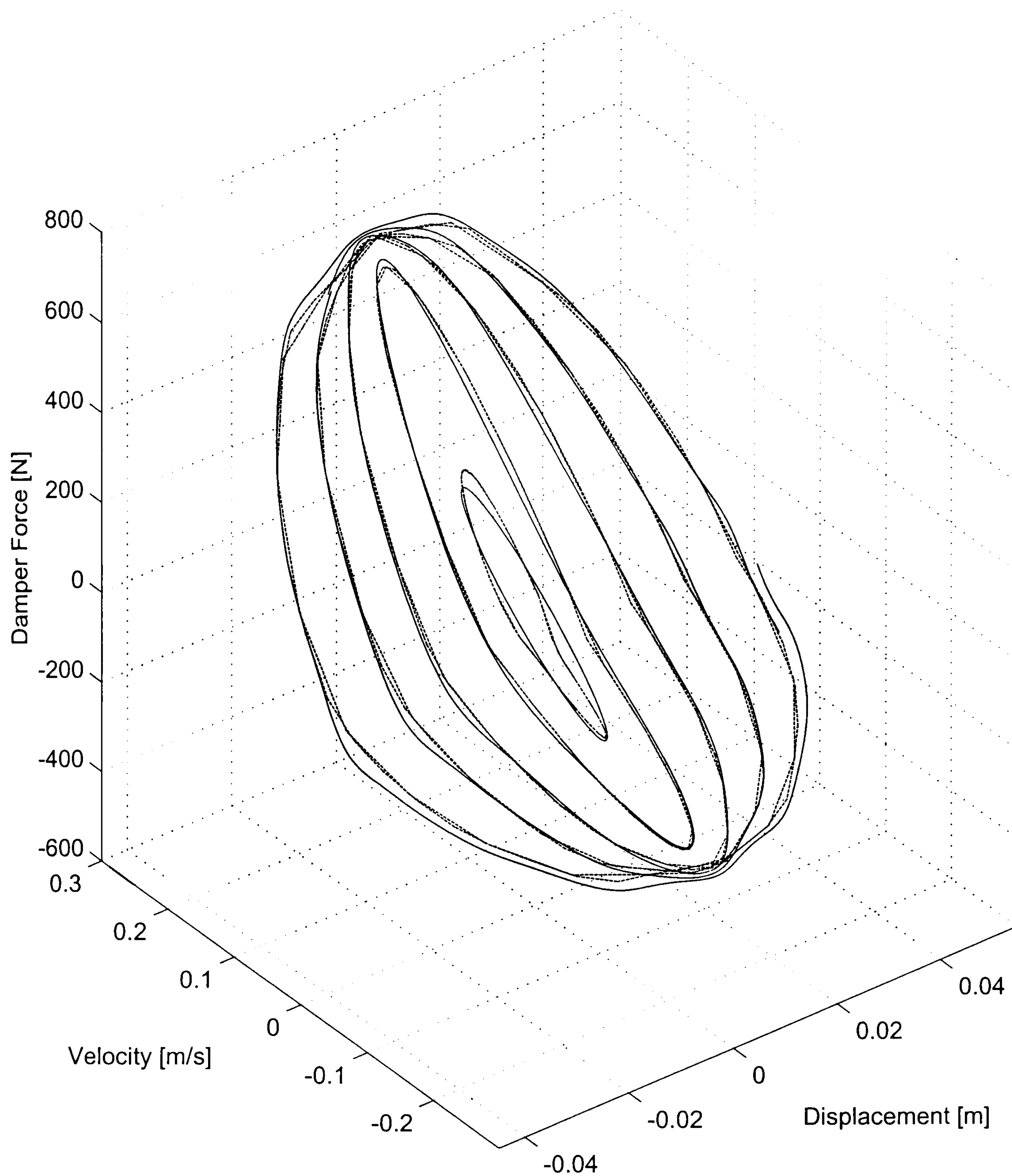


Fig. 3.4.4 Three-dimensional sectioned plot of fitting result obtained from Model 3;
 - measured, -- modelled.

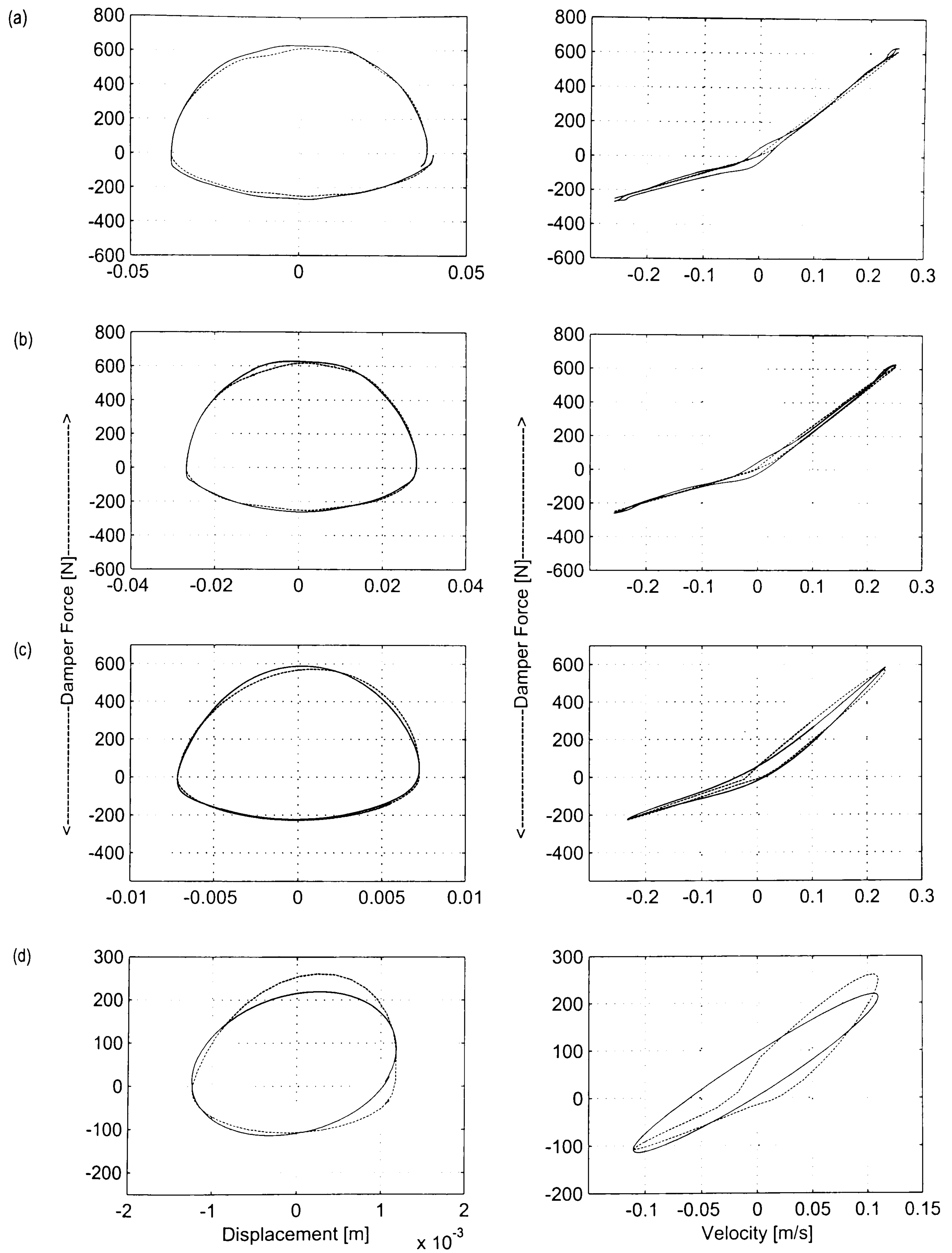


Fig. 3.4.5 Section plots of fitting result obtained from Model 3;
- measured, -- modelled.

The compliance force in the spring element is related to the damping force produced by the bilinear dashpot element according to the equation (3.3)-(3.5) given in Table 3.3.1. Since the spring element is linear, the evidence seen on the force-velocity plots is that the hysteresis loop in the compression direction is smaller than that in the rebound direction. When trying to predict hysteresis loop and asymmetrical behaviour, a larger model error is seen at high frequencies, see part (d) of the figure. Though the asymmetrical characteristic and hysteresis loop were dealt with, the overall model error is accumulated from those due to exclusion of the friction and the ‘not’ perfect treatment of the hysteresis and nonlinear damping characteristic. Having discussed the model errors, its performance however is a satisfactory and justifiable one, as it requires small modelling effort and offers a fast and relatively good quality of fit.

Two further modifications were made to Model 3 in an attempt to improve the quality of fit. First, in stead of linear spring element, a cubic spring is used in Model 4. Second, further from Model 4, a transition factor for the damping rates was incorporated in Model 5 to reduce a sudden change of damping rate, hence the forces, between compression and rebound directions. The resultant model is similar to that of Besinger et al [115], with five parameters to be estimated; two damping rates, two stiffnesses for the spring element, and the damping rate transition factor. The four section plots of the damper force against displacement and velocity obtained from Model 4 are very similar to that of Model 3. In fact, the quality of fit in term of MSE value is 0.35%. A small amount of accuracy is gained for the modelling effort. The time spent in the estimation procedure for Model 4 is more than twice as long as those

of Model 3. Similar arguments can be drawn when comparing the results from Model 5 against that of Model 3. The lowest MSE of 0.31% obtained from Model 5 is slightly lower than that of Model 4, as expected since Model 5 is more complex and contains more parameters than Model 4. However the mean value from 50 independent runs is 0.36%. This is because of the difficulty due to local minima encountered during the parameter search. The time spent in the parameter estimation for Model 5 is more than five times of that of Model 3. The gain in quality of fit obtained by increasing number of parameter and/or increase complexity of the model in Model 4 and 5 respectively is not as much as the time lost during the parameter search.

3.4.2.2 Numerical Search Methods: Performance Comparison.

The best parameters for each model were identified using four numerical search methods; the GB, DS, DE and dvHDE methods. The results were gathered from 50 independent runs, and are presented in Table 3.4.1 and Table 3.4.2. The comparison is done by considering the mean, m , and the standard deviation, σ , of the final cost function, denoted J_{MSE} , the CPU-time, denoted T_{CPU} , and the number of function evaluations, N_{feval} . The performances of the four methods are discussed one by one below.

	GB Method						DS Method					
	Mean Values			Standard Deviations			Mean Values			Standard Deviation		
	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$m_{N_{feval}}$	$\sigma_{J_{MSE}}$	$\sigma_{T_{CPU}}$ [s]	$\sigma_{N_{feval}}$	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$m_{N_{feval}}$	$\sigma_{J_{MSE}}$	$\sigma_{T_{CPU}}$ [s]	$\sigma_{N_{feval}}$
Model 1	18.966	2.34	25.1	0.000	0.85	7.7	18.966	9.05	49.6	0.000	0.53	4.4
Model 2	1.256	11.48	76.4	0.000	5.65	35.4	1.256	30.16	159.5	0.000	4.01	20.9
Model 3	1.010	115.39	187.1	0.648	68.24	113.9	0.536	212.75	348.7	0.000	30.88	51.2
Model 4	0.396	445.51	553.1	0.051	154.87	181.4	0.481	492.97	573.5	0.404	338.26	388.3
Model 5	1.459	407.42	398.7	0.995	162.90	161.3	1.960	1518.05	1542.6	1.995	405.58	434.1

Table 3.4.1 Estimation results from the GB and DS methods.

Note:

Model 1: linear dashpot

Model 2: bilinear dashpot

Model 3: linear spring in series with bilinear dashpot

Model 4: cubic spring in series with bilinear dashpot

Model 5: cubic spring in series with bilinear dashpot (with transition factor)

	DE Method						dvHDE Method					
	Mean Values			Standard Deviations			Mean Values			Standard Deviation		
	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$m_{N_{feval}}$	$\sigma_{J_{MSE}}$	$\sigma_{T_{CPU}}$ [s]	$\sigma_{N_{feval}}$	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$m_{N_{feval}}$	$\sigma_{J_{MSE}}$	$\sigma_{T_{CPU}}$ [s]	$\sigma_{N_{feval}}$
Model 1	18.966	20.23	114.4	0.001	8.31	47.1	18.966	8.97	54.6	0.000	2.36	14.4
Model 2	1.256	31.32	182.6	0.000	13.45	79.9	1.256	23.76	128.8	0.000	4.59	24.7
Model 3	0.536	172.45	319.5	0.001	86.23	145.5	0.536	121.43	193.8	0.000	22.90	38.4
Model 4	0.353	460.42	544.2	0.001	196.48	229.4	0.353	332.2	378.2	0.001	103.30	119.9
Model 5	0.915	946.68	810.6	0.685	75.43	66.1	0.360	667.27	630.9	0.011	46.50	27.9

Table 3.4.2 Estimation results from the DE and dvHDE methods.

Note:

Model 1: linear dashpot

Model 2: bilinear dashpot

Model 3: linear spring in series with bilinear dashpot

Model 4: cubic spring in series with bilinear dashpot

Model 5: cubic spring in series with bilinear dashpot (with transition factor)

Model 1 is a simple one-dimensional optimisation problem that all four methods considered were able to find the optimum point. The GB method performed best among the four. The method made use of the gradient information of the cost function effectively, thus moved quickly from any randomly generated initial point to the optimum point. This is evident from the results in Table 3.4.1 that the GB required the least number of function evaluations, hence minimum amount of time. The other three methods, DS, DE and dvHDE, are direct search methods that do not have the information on the gradient of the cost function. They all required greater number of function evaluations, therefore more time, when compared to the GB method. The DE method is the least favourable. Its accumulated time comes from its inherent parallel mechanism in the search procedure, as it is a population-based method. The dvHDE is also a population-based method and has a parallel search mechanism. However, the number of function evaluations is less than that of the DE method. The dvHDE was able to reduce the number of function evaluation, hence reduce the CPU-time. The results presented for the dvHDE method in Table 3.4.2, only Model 1, was obtained without incorporating the acceleration and migration operations. The performance of the dvHDE method has been shown to be similar to that of the DS method. The dvHDE is slightly better, justified from similar CPU-time and number of function evaluations, but with a larger standard deviation of CPU-time. Nevertheless, the results from Model 1 have shown the use and advantage of the dvHDE over the original DE method, though it is not able to compete with the GB method for this one-dimensional optimisation problem.

Model 2 where the damper was modelled using a bilinear dashpot element is a two-dimensional problem. Similar arguments as that from Model 1 can be said for Model 2. All methods were able to find the optimum point every run. The GB method has shown to be the best method, with the least number of function evaluation and time. The dvHDE method, with the acceleration and migration incorporated, performed better over both the DS and DE methods. Comparing the dvHDE against the DE method, the dvHDE method required less number of function evaluations to find the optimum point. The amount of time required by the dvHDE is better than the DE method by approximately 24%.

In Model 3, the damper was modelled as a linear spring in series with a bilinear dashpot element. The parameter estimation is a three-dimensional optimisation problem. Complexity of the problem increases not only in term of number of the parameters to be identified, but also numerical solution when solving the dynamic equations of the model. To obtain the modelled damper force from Model 3, numerical integration/differentiation was required in order to solve for the additional variables, x_s and \dot{x}_s . Apart from being able to find the optimum point consistently, it would then be further advantage if a parameter search method performs evaluation of the cost function at a minimum number of times, so that the overall end result is the problem solved quickly.

With the GB method, the final fitness value, MSE value, was different from run to run. The mean and standard deviation values of the MSE were the highest among the four methods, see Table 3.4.1. The method more often failed to obtain the optimum value than it succeeded to do so. Though, the average CPU-time was less than that of

the other methods, it rather suggested that the algorithm termination criteria was met and the search was stopped in relatively short time, regardless of whether it was the optimum point or not. It therefore can be concluded that the GB method encountered problem of obtaining local minima and failing to escape from them, hence unable to consistently find the optimum point. Its ability to find the optimum point for Model 3 depended on the initial starting points. Overall performance was the least favourite.

The DS, DE and dvHDE methods were able to find the optimum point every time.

The DS method had a slow convergence rate, as it took a longer time and more number of function evaluations, see Table 3.4.1. The DE and dvHDE performed better than the DS method. The dvHDE method performed the least number of function evaluations, hence reduced the time spent by 29.6% when compared to the DE method.

Fig. 3.4.6 (a) and (b) show, respectively, typical plots of MSE value belonging to the best chromosomes and the population diversity measured at each iteration obtained from the dvHDE method. The same plots for the DE method are presented in Fig. 3.4.7. The corresponding plots of the estimated parameters are given in Fig. 3.4.8 for the dvHDE method and Fig. 3.4.9 for the DE method. These plots give some information about the behaviour of the two methods. For ease of observation toward the search termination, the iteration-axis is shown on a logarithmic scale.

Both the DE and dvHDE methods quickly evolve the population toward a better search area, as suggested by the reduction in MSE values in parts (a) of Fig. 3.4.6 and Fig. 3.4.7.

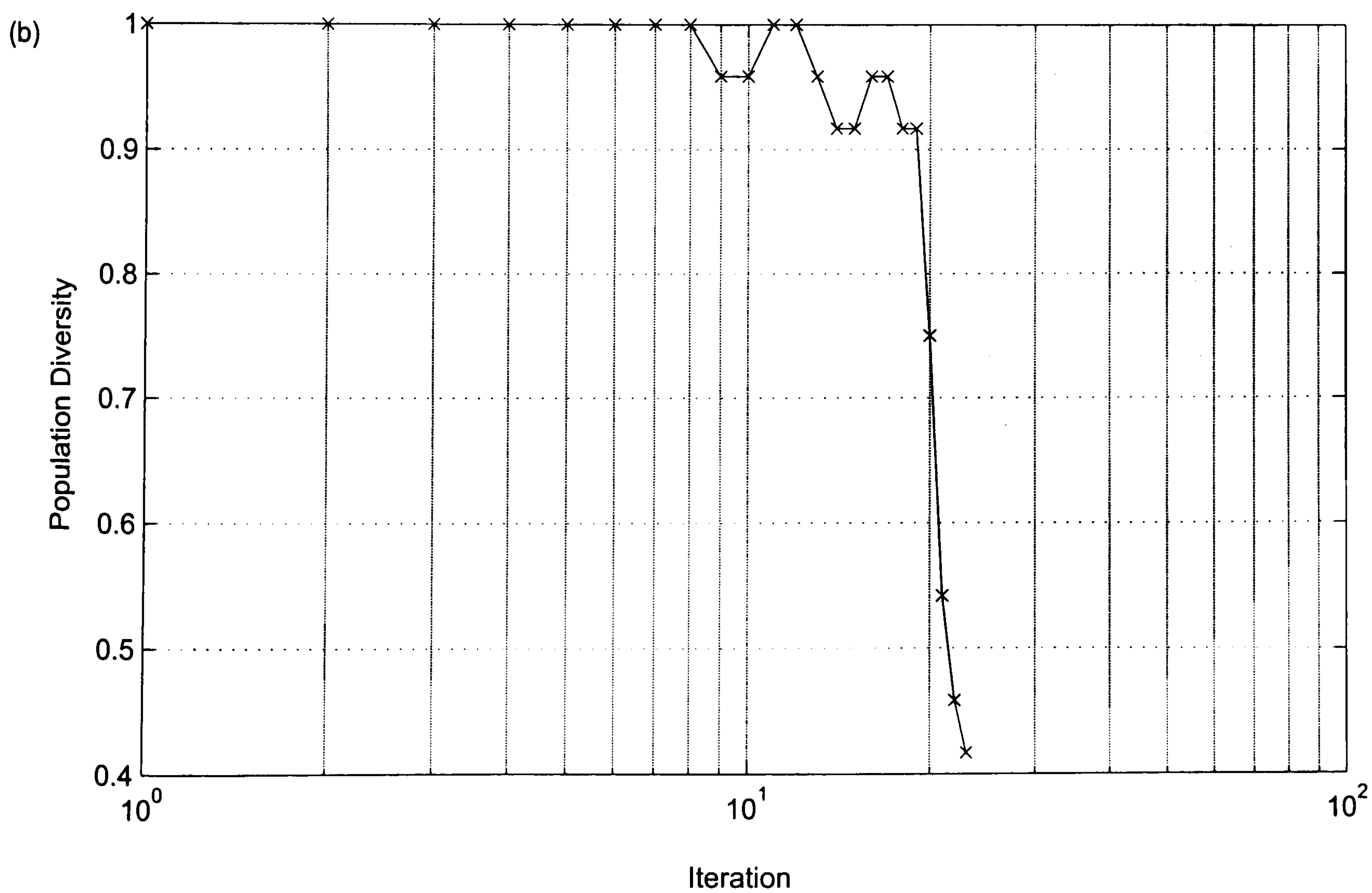
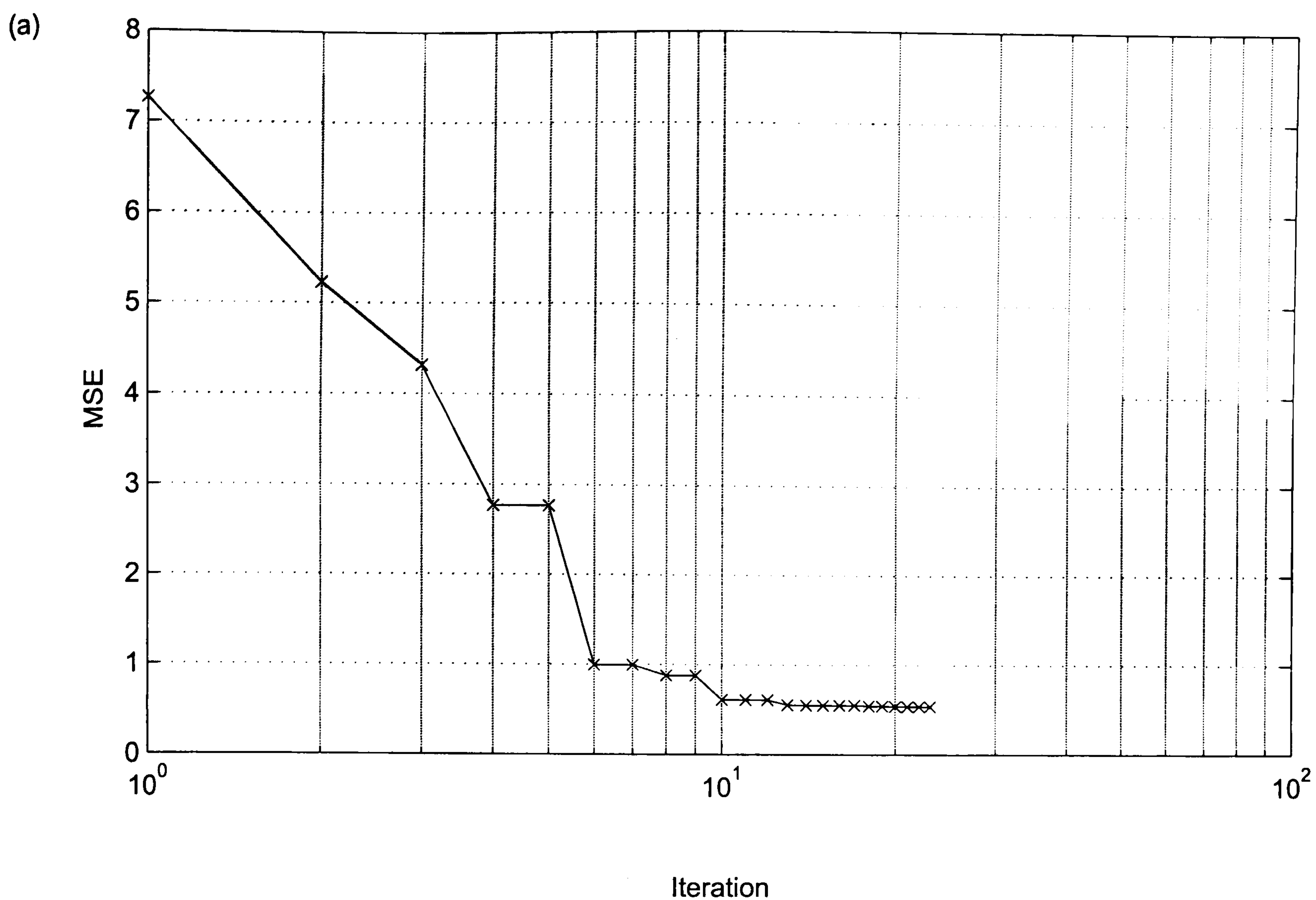


Fig. 3.4.6 Typical plots of (a) best MSE values and (b) population diversity obtained from the dvHDE method.

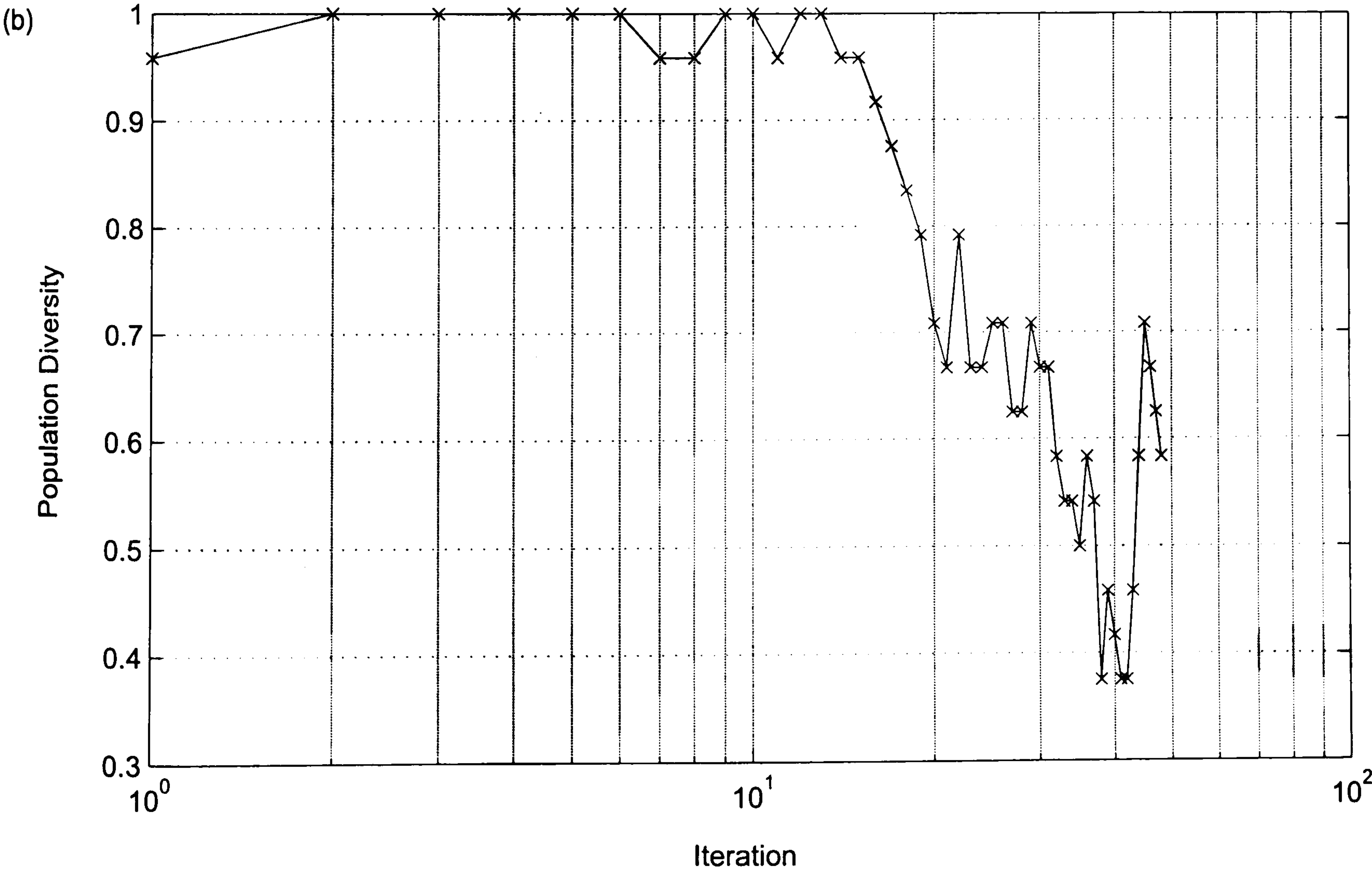
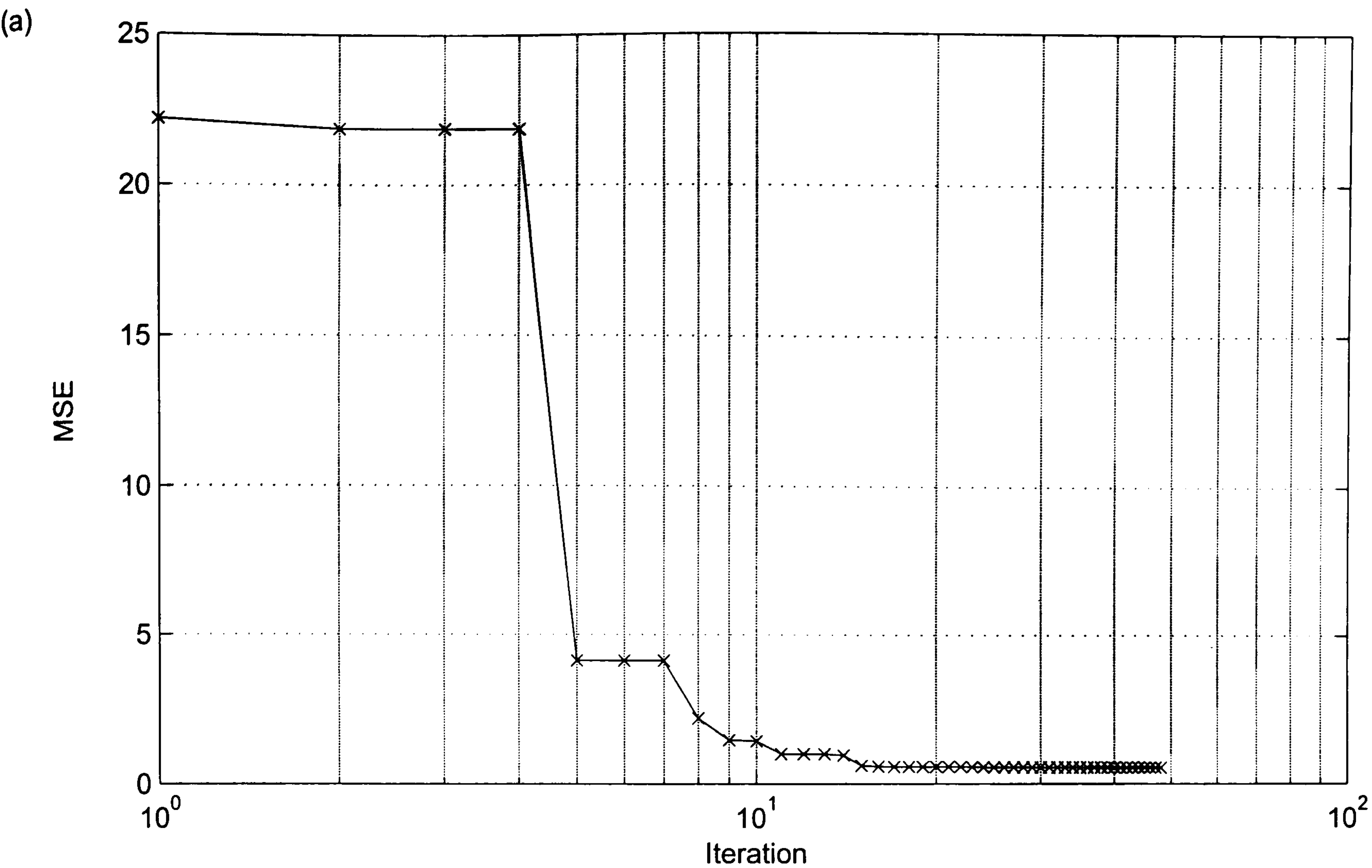


Fig. 3.4.7 Typical plots of (a) best MSE values and (b) population diversity obtained from the DE method.

The population of parameters was updated by the selection operation, so better individuals replace the worse ones. In the early generations, where the population was diverse, the value of population diversity was, close or equal to, one, see parts (b). As evolution progressed, the value gradually reduced as more and more individuals were close to the best one and/or close to each other. Another important point need be mentioned, is that there are several iterations that the DE method failed to update the MSE value, hence the best individual, or only a small reduction in the MSE value was achieved. This is also evident from the plots of estimated parameters, when all three parameters remained unchanged, see Fig. 3.4.9. Comparing Fig. 3.4.7 and 3.4.9 with those of the dvHDE method in Fig. 3.4.6 and Fig. 3.4.8, the latter shows a better convergence rate. The choice of using the acceleration and migration are optional. This allows for more ‘tuning’ in addition to the algorithm control parameters, such as size of population and mutation and cross over factors in the original DE method. With proper acceleration and migration operations included, the dvHDE was able to reduce the number of function evaluation and move the search space quicker toward the optimum point.

Model 4, where the damper was modelled using a cubic compliance spring in series with a bilinear dashpot, which was a modified version of Model 3 by further including one additional parameter. As discussed earlier, the quality of fit increases, as shown in Table 3.4.2 from MSE value of about 0.5% to 0.35%. The complexity of the optimisation problem increases from three-dimensional to four-dimensional one. The time required to find the optimum point increases by more than twice that of Model 3.

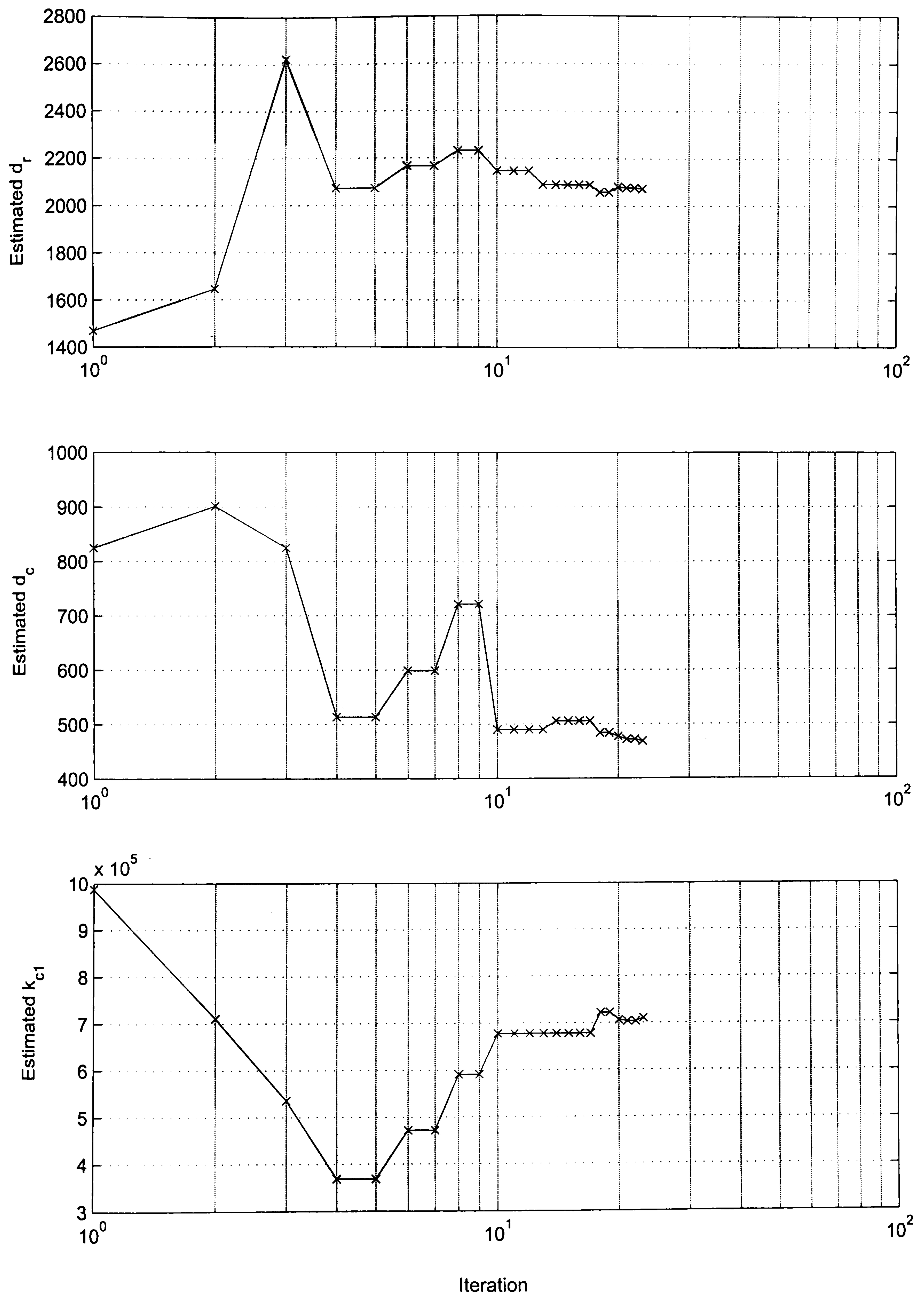


Fig. 3.4.8 Estimated parameters of Model 3 using the dvHDE method.

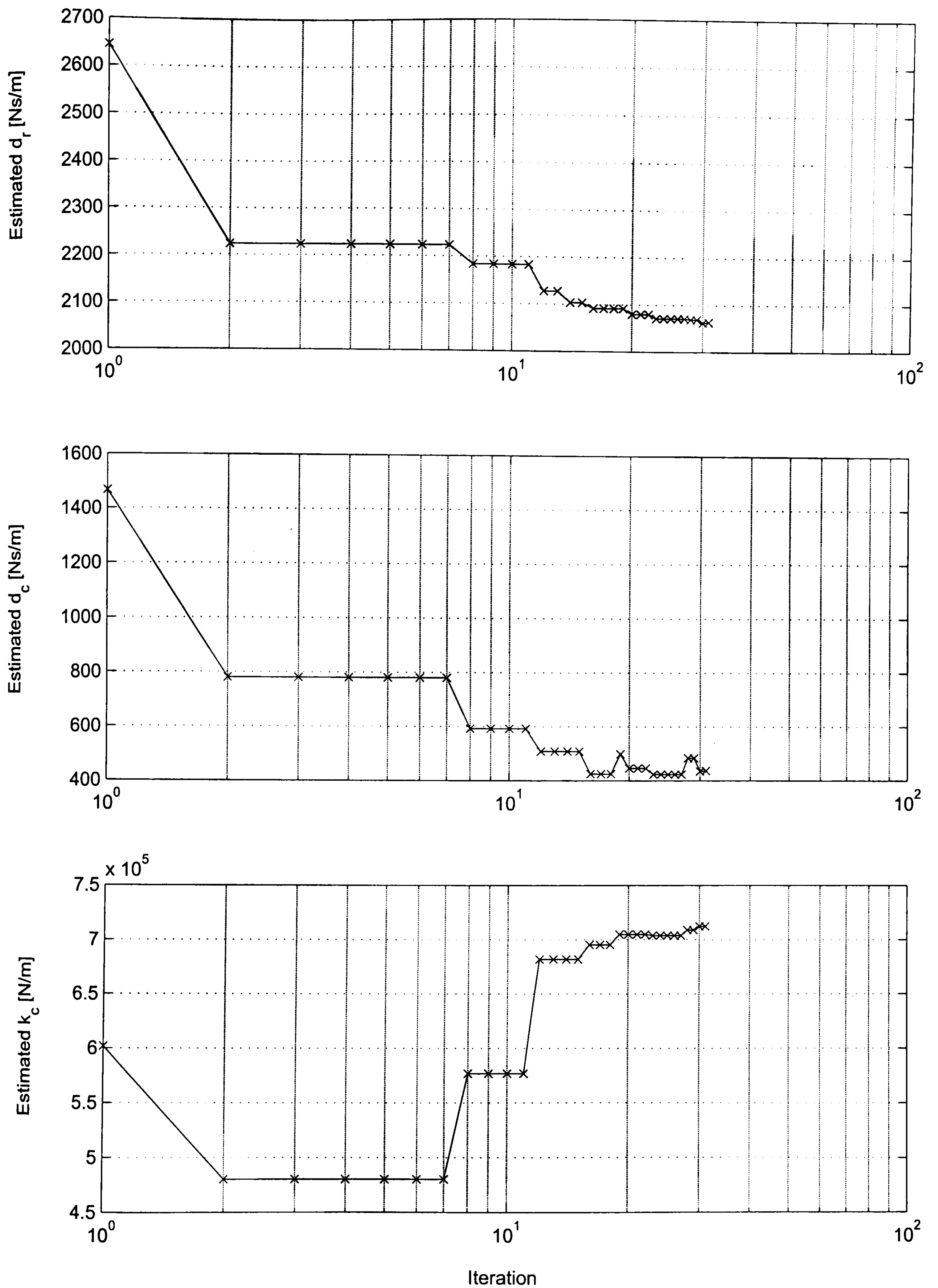


Fig. 3.4.9 Estimated parameters of Model 3 using the DE method.

The GB again had difficulty in finding the optimum point due to local minima, as evident from the average value of the final MSE and, the problem can be said to be a consistent one. This is indicated by a high mean value of the CPU-time, with relatively small standard deviation. The DS method was able to find the optimum point, however, occasionally the search was trapped in one or two local minima. The average value of the final MSE is 0.48%, thus greater than its 'best' values 0.35%. When the optimum point was found by the DS method, it has a slow convergence rate on average. The time it spent depended on the initial starting points, as suggested by a high the standard deviation of CPU-time. While the GB was consistently trapped in local minima, the DS method suffered one or two local minima, and when not trapped, it was shown to have a slow convergence rate. The DE and dvHDE methods both were able to find the optimum point every run. In term of time, the dvHDE method was shown to be better than the DE method by almost 28%. It performed the least number of function evaluations. Also the acceleration operation improved the convergence rate. These two capabilities, apart from having a parallel mechanism, were the main features that made the dvHDE method the most favourable choice for the problem studied.

In an attempt to further improve the quality of fit, Model 4 was modified, and the resultant model was referred to as Model 5. A sudden change in damping rates between those of the compression and rebound directions considered so far in Model 2-Model 4 was not true physically. The damping rather changed continuously between the two phases (i.e. compression and rebound). Model 5 took into account of this fact and formulated the damping rate as a smooth function of the velocity using

the transition factor α . The inclusion of a cubic compliance spring has shown to improve the quality of fit. This was also incorporated into Model 5. The model thus contained five parameters; two damping rates, two stiffnesses, and one transition factor. Again, the complexity of the optimisation problem increased not also in term of number of parameters, but also dynamic behaviour of the model. A longer time was expected and shown to be the case, see T_{CPU} in Table 3.4.1 and 3.4.2. Apart from the GB method, which suffered even more from the same problem as with Model 3 and 4, the time required spent for all methods increased. For the DS method, similar arguments to Model 4 can also be drawn, however, the problem became worse. The method got trapped in several places in the search space, when trapped, the algorithm kept on searching until the termination criteria was met. This registered a high value of CPU time spent. When not trapped in local minima, the method still required a long time to find the optimum point. This suggests again that it has a low convergence rate. The DE method required less time than the DS method, but more than the dvHDE method. The DE method was able to find the optimum point, but often got trapped in local minima of two or three places in the search space, as suggested by a high value of σ_{MSE} . Fig 3.4.10 (a) shows a typical plot of MSE value belonging to the best chromosome at each generation. The corresponding population diversity measured is presented in Fig. 3.4.10 (b).

Again it is observed from part (a) of the plot that the MSE value remained unchanged for several consecutive iterations. After one point in the search process, the population diversity rapidly reduced, seen on the plot from 1.0 to 0.8. After that, the reduction in MSE value was very small or zero. The population diversity can be said to be unchanged, though small variations were observed on the plot.

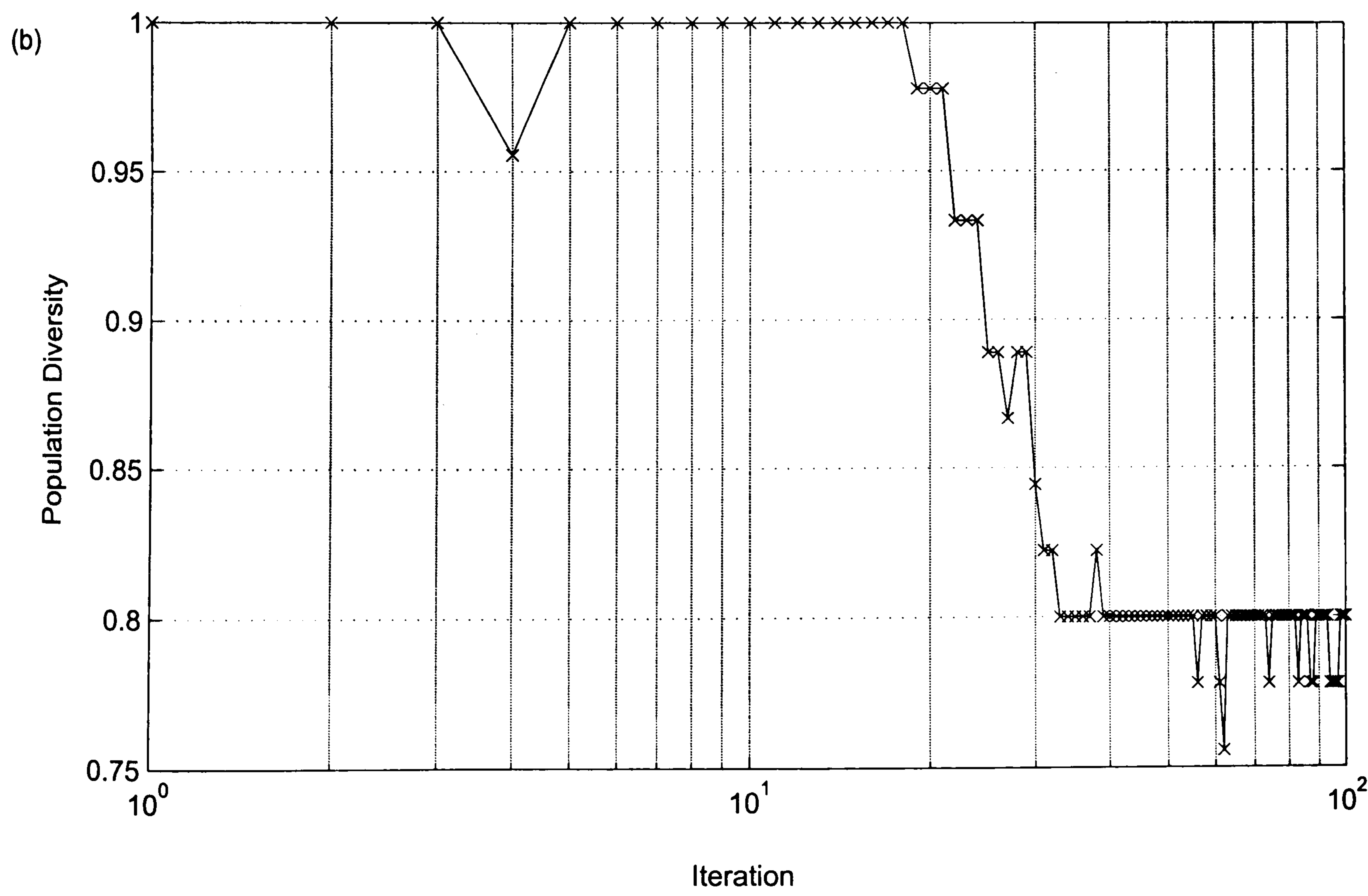
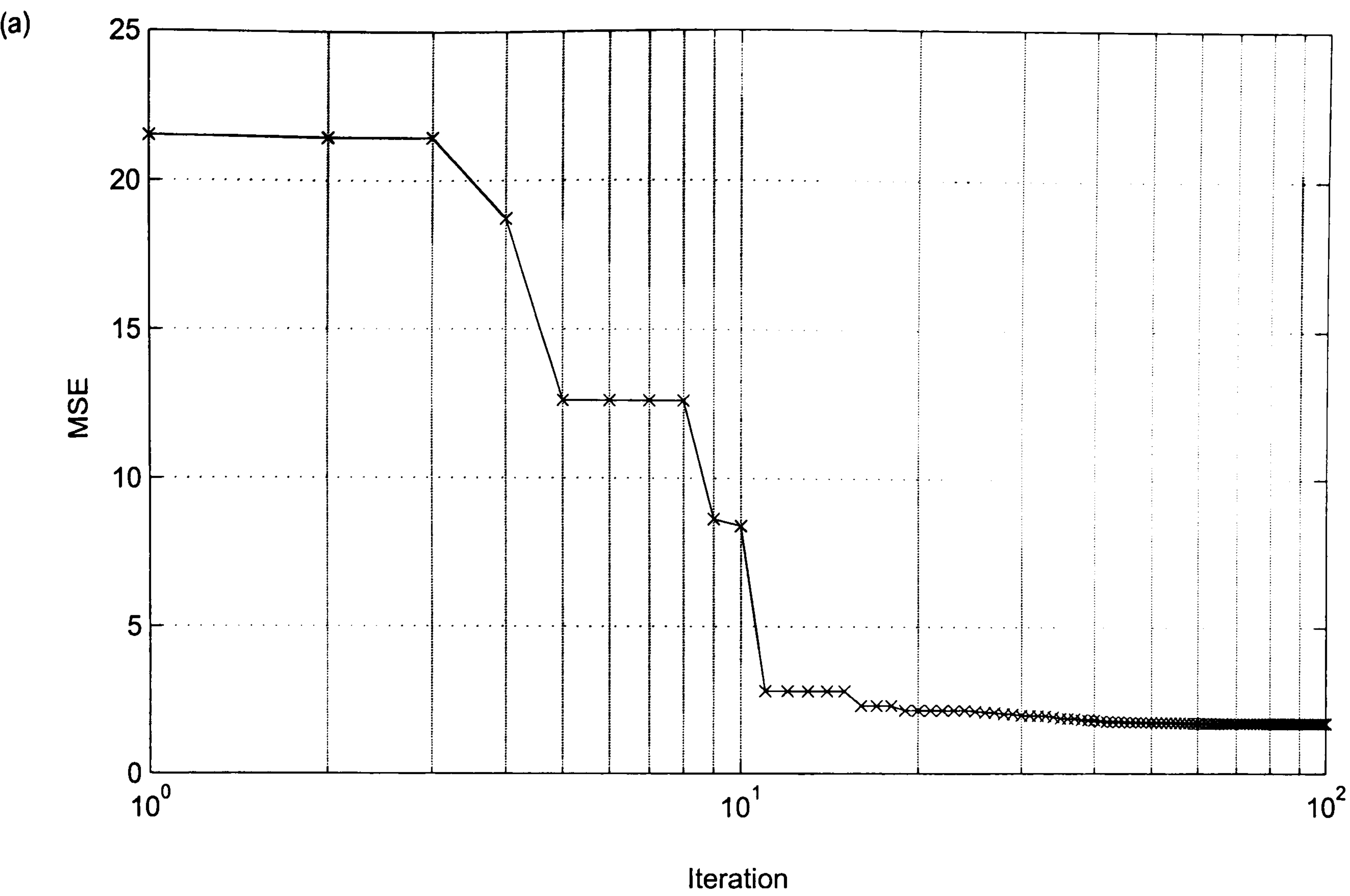


Fig. 3.4.10 Typical plots of (a) best MSE values and (b) population diversity obtained from Model 5 using the DE method.

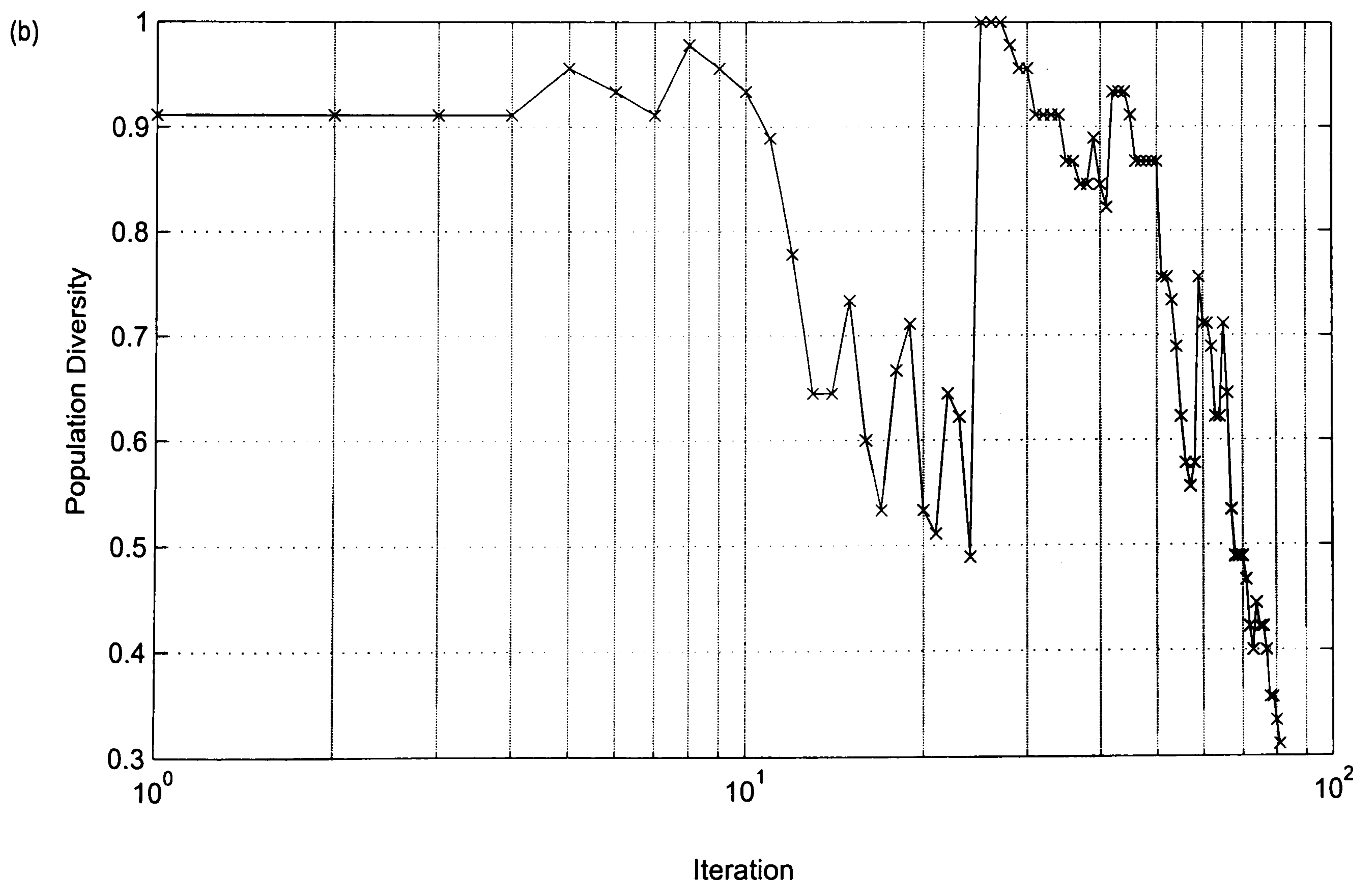
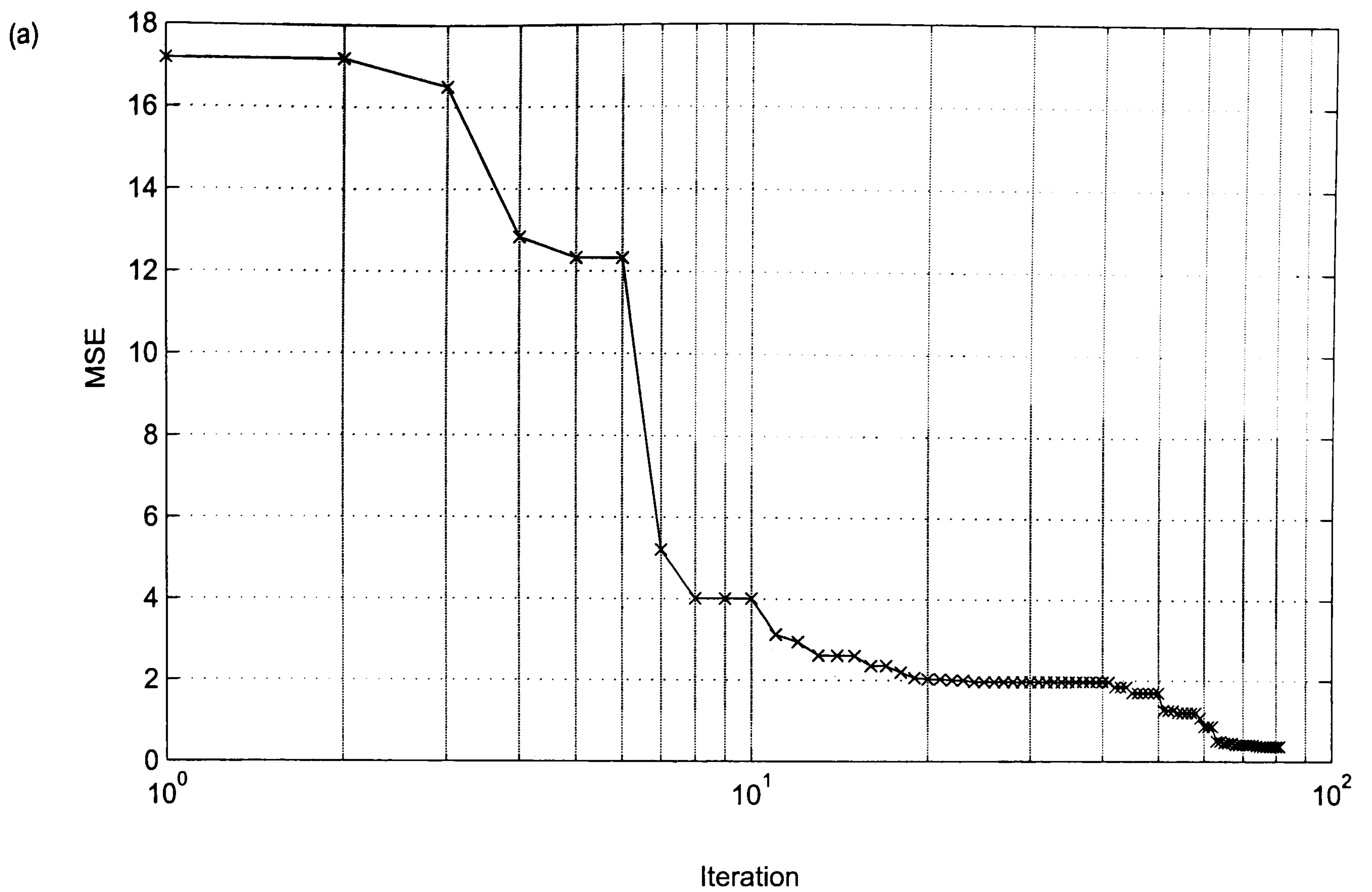


Fig. 3.4.11 Typical plots of (a) best MSE values and (b) population diversity obtained from Model 5 using the dvHDE method.

The circumstance here suggests that the mutation and crossover operations, though were successful in producing better individuals that replace the worse ones in the current generation, as evident from variation in population diversity, however failed to generate an individual or individuals which could move the search further. The search was then trapped in a local minimum. The DE method employed no mechanisms to make an escape, hence failed to find the optimum point. When comparing with the dvHDE method, see Table 3.4.2, the dvHDE method again performed less number of function evaluations. With careful set-up of the acceleration and migration operations, it was possible to improve the convergence rate and avoid ending at a local minimum point. To give a visual comparison with the DE method, typical plots of MSE value and population diversity obtained from the dvHDE method are presented in Fig. 3.4.11. The MSE value consecutively reduced and at one point where the MSE value was unchanged and the search approached a local minimum, the migration operation was activated. A sudden increase of population diversity being observed in Fig. 3.4.11 (b) when the migration took place. A successful migration operation made an escape and moved the search towards the optimum point. Based on the CPU-time, the dvHDE performed better over the DE method by 29.5 %.

In summary, as far as modelling of the test damper is concerned, several important points can be addressed here. Model 1, linear dashpot, though offered a simple and fast identification, it was however not sufficient to accurately predict the damper dynamic behaviour of the damper. The quality of fit of Model 1 was unacceptable with an MSE value of 19%. The quality of fit was improved significantly with Model 2, where the damper was modelled as a bilinear dashpot element. The model described the dynamic behaviour of the damper better, as indicated by the MSE value

of 1.26%. Apart from being unable to capture the friction and hysteresis, the bilinear character of Model 2 predicted the damper behaviour at low and intermediate frequency intervals relatively well, however with a large error at high frequencies. Model 3, where the damper was modelled as a linear spring in series with a bilinear dashpot, gave a further improved quality of fit with the MSE value of 0.54%. The plots of the measured and modelled force against displacement have good agreement, and the corresponding force-velocity plots showed that the model was capable of capturing the bilinear characteristics and the compliance loop with a good degree of accuracy. The imperfect treatment of the hysteresis, the exclusion of the friction and nonlinear damping characteristic were the main contributors to the model errors. Model 4 and Model 5 were the two further developed versions of Model 3; in stead of a linear compliance spring, a cubic spring was used in Model 4, and, in addition to the cubic compliance spring, a damping rate transition factor was used in Model 5. The qualities of fit of the two models were better than Model 3, with the MSE values of approximately 0.35%. The models were more complex than Model 3 in that it contained a larger number of model parameters and involved a larger amount of numerical calculation, hence the estimation time consumed was longer as expected. In fact, with Model 4 and Model 5 the time required to complete the parameter estimation tasks were, respectively, more than twice and five times as long as those of Model 3. While Model 1 and Model 2 were simple but judged as insufficient to accurately describe the damper behaviour, the gains in quality of fit by the additional modelling effort in Model 4 and Model 5 did not balance with a higher price paid in the estimation procedure. It can therefore be concluded that, as far as the modelling accuracy and the estimation time required are concerned, Model 3 has been shown to be the best among the models considered.

The parameter values of all the damper models considered in the investigation were estimated using four different numerical search methods. The performance of the four methods have been analysed and compared, an important summary can be made as follows. First, when optimisation problem is not complex, the standard gradient-based (GB) method is generally better than direct search and population-based methods. Gradient-based methods effectively use the information about the gradient of the cost function with respect to the parameters to make a step of appropriate size moving from a point at the present iteration to the next one. This has been shown to be the case for the estimation of the damper model parameters with linear and bilinear dashpot models, in which the number of parameters were one and two respectively. However, when the search space contains local minima, and is not uni-modal, the gradient-based methods can suffer from being trapped in local minima, and fails to find the optimum point. This problem seems to be encountered more often as complexity of optimisation problem increases. The parameter estimation carried out for Model 3, Model 4 and Model 5 have proved to be examples of this. The DS method, which is a direct search method, has shown to have a slow convergence rate for the problem studied. One reason may be because it is a single point search, not a parallel, population-based such as the DE and dvHDE methods. Its ability and the time required to find the optimum point depend on the initial starting point. The DS method is occasionally unable to find the optimum point as it is trapped in local minima, for examples in Model 4 and Model 5. In general the DE and dvHDE methods are useful and more suitable for a complex optimisation problem where local minima exist. The two methods, which require no information about the gradient of the cost function, are alternatives when calculation of the cost function and its gradient information are computationally expensive or impossible. In the case of

Model 3 - Model 5, there existed several local minima, and the calculation of the cost function for Model 5 was the most time consuming of all. The results from this investigation has shown that the DE and dvHDE method performed better over the GB and DS methods for Model 3 - Model 5. The DE method relied on two main mechanisms, the mutation and crossover operations. However, it has been illustrated in Model 5 that the two mechanisms did failed, and the search was trapped in local minima. In addition, as often evident from the ‘discontinuous’ characteristic on the analysis plots, that there are several consecutive iterations on which the DE failed or did not effectively move the search to a better region. The convergence speed of the DE method was slower than the dvHDE method for all the models considered. The problems of local minima and mis-convergence were dealt with and convergence speed was improved in the dvHDE method. Based on the performances from Model 3 – Model 5, the dvHDE method consistently found the optimum points, and the convergence speed was increased by almost 30%. As far as the performance of the parameter search methods is concerned, the dvHDE method has shown to be the most preferable method.

3.5 Conclusions

In this chapter, system identification and parameter estimation problem of an automotive damper has been presented. Five models, which were increasing in complexity, were considered in the selection of the best model for the damper under test. The set of models considered were derived from combinations of idealised spring

and dashpot elements, which represented important physical properties such as compliance and damping over the frequency range of interest 0.5-30 Hz.

The parameter estimation of each model was formulated as a single objective minimisation of the difference between the measured and modelled damper forces. The performance of one particular model was discussed and compared to the others in terms of the quality of fit to the experimental data and the modelling effort required. Model 1 and Model 2, where the damper was modelled as linear and bilinear viscous damping elements respectively, were simple and gave a fast identification. However, they were judged as insufficient to accurately describe the dynamic behaviour of the damper under test. Model 3, where the damper was modelled as a linear spring connecting in series with a bilinear viscous damping element, was able to capture the dominant dynamic behaviour of the test damper, the bilinear characteristic and compliance. The model gave a quality of fit in term of MSE value of 0.54% indicating a good agreement between the experimental and modelled damper forces. Model 4 (a cubic spring connecting in series with a bilinear dashpot) and Model 5 (a cubic spring connecting in series with a bilinear dashpot, plus a damping rate transition factor) required, respectively, more than twice and five times as long as those of Model 3 to complete the parameter estimation task. The gains in quality of fit by the additional modelling effort in Model 4 and Model 5 were less than the time lost in the estimation procedure. The results have therefore shown that Model 3 was the best among the models considered as it gave the best compromise between the modelling accuracy and the estimation time required.

The best parameters for each model were identified using four different numerical search methods, the GB, DS, DE and dvHDE methods. The performances of the four methods were discussed and compared to one another by considering the mean and standard deviation of the final cost function, the number of function evaluations performed, and the CPU-time from 50 independent runs.

When optimisation problem was not complex such as in the case of Model 1 and Model 2, which were one and two dimensional problems, the GB method effectively made use of the gradient information of the cost function, it thus performed better than the direct search (the DS method) and population-based methods (the DE and dvHDE methods). In Model 3-Model 5, the problem complexity increased not only due to increase in number of parameters but also due to increase in amount of numerical calculation involved. The GB performed the worst among the four methods considered, as it experienced the problem of obtaining local minima, and the GB's performance became worse as the complexity of optimisation increased. The DS was shown to have the worst convergence speed for all the models considered, and was occasionally trapped in local minima. The DE and dvHDE methods were shown to be useful and more suitable for complex optimisation problems. In addition to the 'discontinuous' descending characteristic of the DE method, its working mechanism did fail, as evident in the case of Model 5. The convergence properties, the consistency and speed, of the DE method were worse than the dvHDE method for all five models considered. Based on the performance from Model 3-Model 5, the dvHDE method was shown to consistently found the optimum with convergence speed better than the DS and DE method by approximately 44% and 30% respectively. The dvHDE was thus the best among the method considered.

CHAPTER 4

PARAMETER ESTIMATION OF A SINGLE WHEEL STATION

4.1 Introduction

A wheeled vehicle represents a complex system with many degrees of freedom. A system that simplifies the complex dynamic behaviour of the whole vehicle is of important benefit because it allows the investigator to concentrate on a limited range of its behaviour. A ‘single wheel station’ test rig is such a simplified system that represents the vertical dynamics of a quarter of a vehicle. The rig consists of two masses representing the body and axle between which the suspension spring and damper are installed. Further simplifications are made by replacing the tyre by a coil spring, which assumes no damping in the tyre, and suspension linkages are excluded all together. The rig allows different set-ups for vehicle suspension investigation to be carried out, for example the following can be easily varied; mass of the body and axle, tyre stiffness, suspension stiffness and damper. In the development of a parameter estimation for a full vehicle, the single wheel station provides an ideal system to test the identification and parameter estimation method.

In this chapter, system identification and parameter estimation procedures are exercised to obtain a model and its parameters for the single wheel station. The parameters of the single wheel station represent an independent suspension of a medium sized family car. The detailed description of the test rig will be given later in the chapter. The task here is again to search for a model, with a minimal number of parameters, which gives an acceptable accuracy in describing the dynamic behaviour

of the system in the frequency range of interest. The model is intended to be used for vehicle engineering purposes such as in fault diagnosis application and simulation study such that the model parameters represents meaningful physical properties. A white-box model, where the equations of motion and physical parameters are used in the model building, is thus here preferable to black-box models.

This investigation is another example aiming at demonstrate the use of the newly developed dvHDE method to real-world applications. The investigation carried out involves both parameter estimation in the time and frequency domains. The estimation has been formulated as a multi-objective optimisation problem, where only easy-to-measure system outputs are needed for the optimisation procedure. Comparing this with the parameter estimation of the automotive dampers in the last chapter, the complexity of the problem increases both due to greater number of parameters to be identified and more complex behaviour of the system being studied. The investigation presents and discusses the performance of different numerical search methods when problems occur in their operation.

The chapter is presented as follows. The single wheel station is modelled based on a quarter vehicle model, which can be derived from a full vehicle model by making certain assumptions. The derivation of the quarter vehicle model is therefore discussed first, in section 4.2.1, and the equation of motion (EOM) of a linear quarter vehicle model are also given. The model is then used to study the influence of the parameter variations on its responses in section 4.2.2. This is done by analysing the frequency responses of the model using Bode plots and observing the system pole locations. To obtain the frequency responses from the model when nonlinear elements

are included, a method based on the Fourier series is developed and presented in section 4.2.3. The method is then validated and used later in the parameter estimation in frequency domain. Section 4.3 describes the single wheel station test rig and the instrumentation. A study on the system measurement noise has been carried out to acquire some information regarding its noise properties. The frequency response tests of the single wheel rig were carried out and the results are presented in section 4.3.4. The parameter estimation in the time domain is formulated as minimisation of five system outputs. These signals are easy to measure. They consist of two displacement signals and two acceleration signals from the sprung and unsprung masses, and one signal representing the dynamic tyre force measured by a load cell. For the parameter estimation in the frequency domain, the objective function in the optimisation procedure is formulated from experimentally obtained frequency responses and the corresponding ones from the model using the technique developed in section 4.2.3. The estimation problems for both the time and frequency domains are thus multi-objective optimisation problems. The estimation technique used to tackle the problems is described in section 4.4. The estimation results are then presented in section 4.5. In this section, the quality of fit for the different models as well as the performances of the different parameter search methods are compared and discussed. Finally in section 4.6, conclusions are drawn regarding the model quality in representing and describing the system dynamics, parameter estimation procedure, and the performance of the parameter search methods.

4.2 A Quarter Vehicle Model

4.2.1 Derivation of a quarter vehicle model

In this section the use of a quarter vehicle model (single wheel station) is justified and equations developed as its response investigated. A vehicle is made up of many components distributed within its exterior envelope. It thus represents a complex system with many degrees of freedom. However, a sufficient representative model can be obtained with certain assumptions. If a vehicle is assumed to consist of a rigid body mass having one wheel at each corner, with longitudinally and laterally independent suspensions connecting the body and wheels, a seven degree of freedom (7DOF) vehicle model may be obtained, as shown in Fig. 4.2.1. The seven degrees of freedom are roll, pitch and bounce motions of the vehicle body, and bounce motions of each of the four wheel-corners. The conventions of the vehicle axis system used when formulating the equation of motion and analyzing the vehicle behavior is given in Appendix B.

When the vehicle is travelling over long wavelength components, the coherence between the left and right tracks is likely to be high, while travelling over short wavelength components the interactions between left and right sides is small. This can be used to further reduce the model to one with four degrees of freedom (4DOF) having no width and only two wheels, see Fig 4.2.2(a).

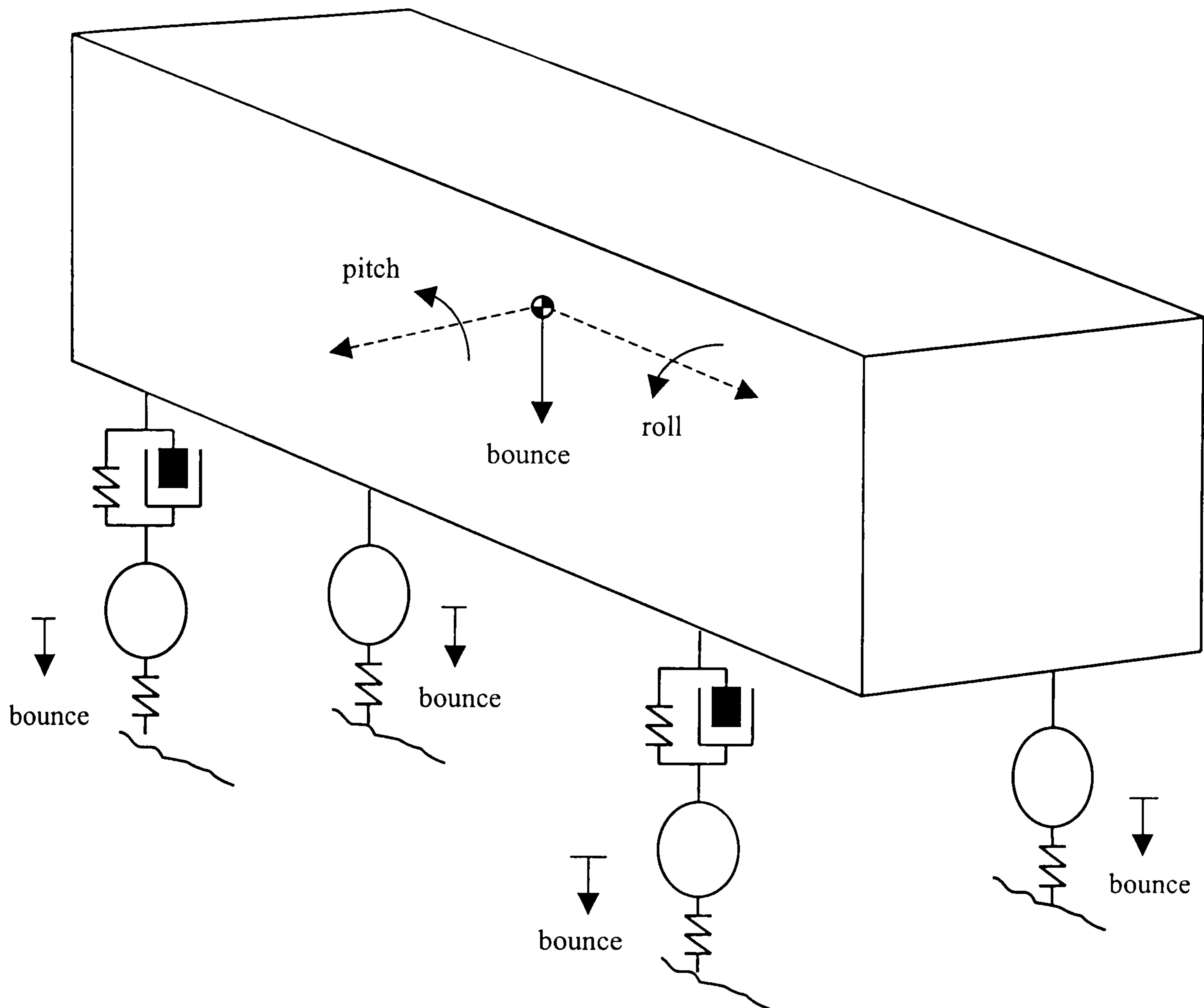


Fig. 4.2.1 A seven degree of freedom (7DOF) vehicle model.

The sprung mass in Fig 4.2.2(a) may be replaced by three point masses rigidly connected by a massless beam, as shown in Fig 4.2.2(b). The two arrangements are dynamically equivalent if the total masses are the same, the mass centers are in the same position, and the pitch moment inertias are equal. Mathematically these conditions are;

$$m_s = m_f + m_c + m_r \quad (4.1)$$

$$m_f l_f = m_r (l_s - l_f) \quad (4.2)$$

$$I_s = m_f (l_f)^2 + m_r (l_s - l_f)^2 \quad (4.3)$$

where

I_s = the pitch moment of inertia

m_s = the sprung mass

m_f = the portion of the sprung mass at the front suspension

m_r = the portion of the sprung mass at the rear suspension

m_c = the portion of the sprung mass at its center

l_f = the longitudinal distances from the sprung mass center to the front axle

l_s = the vehicle wheelbase

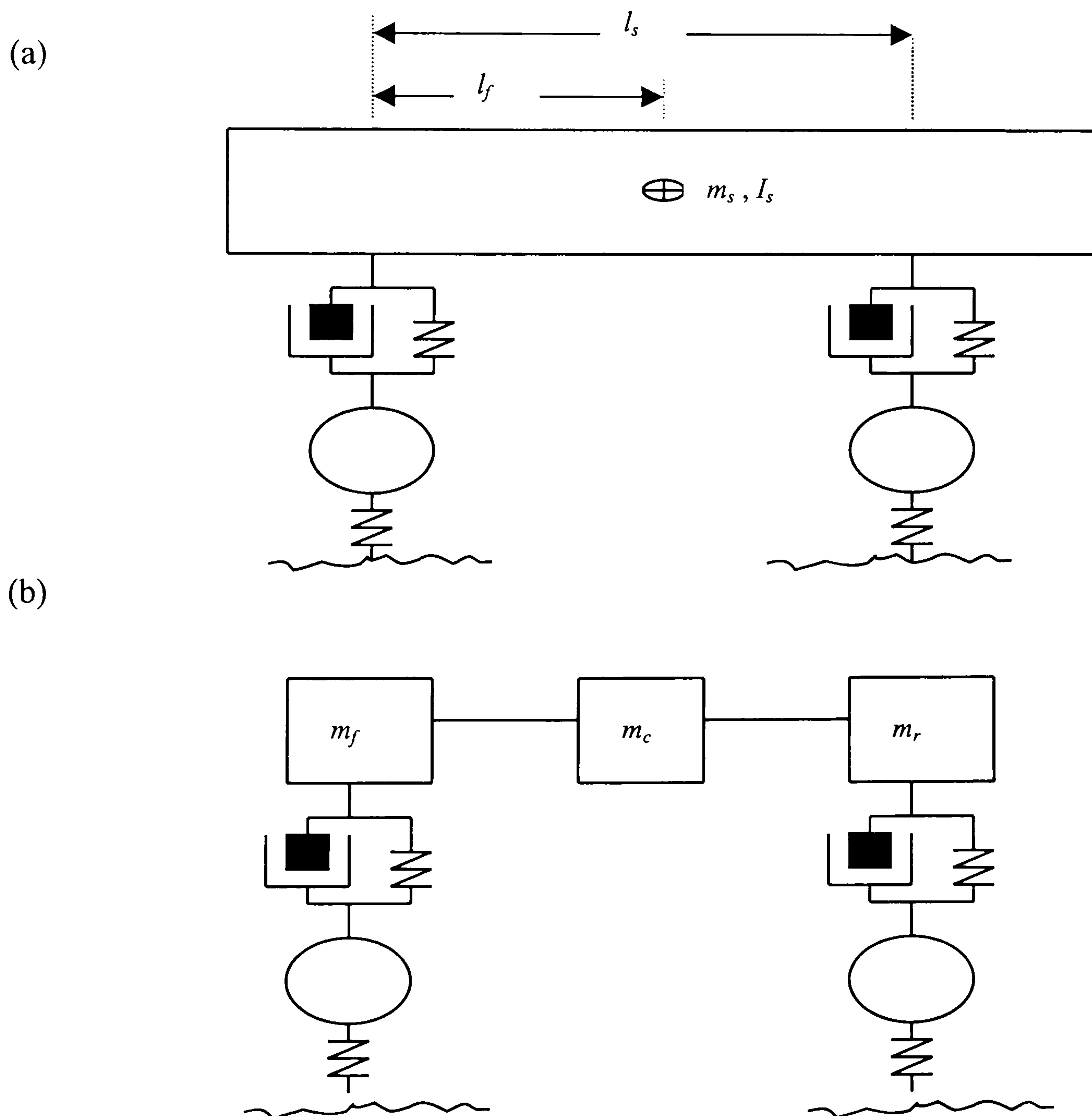


Fig. 4.2.2 A four-degree-of-freedom (4DOF) vehicle model (a), and dynamically equivalent model (b).

If it is assumed that a mass distribution in m_s is such that

$$I_s = m_s l_f (l_s - l_f) \quad (4.4)$$

which is approximately the case for many cars [117-118], combining equations (4.1) - (4.4) yields $m_c = 0$. In this case interactions between the front and rear do not occur. The vehicle model is then reduced to two 'sub-models', with the motion of m_f determining the behavior of the front suspension, the motion of m_r for the rear suspension behavior, and the motion of points within the wheelbase may be determined by geometric proportioning. For the front or rear suspension, the resultant model has therefore two degrees of freedom (2DOF), also known as a quarter vehicle model, shown in Fig. 4.2.3. The quarter vehicle model is widely used representation of a vehicle suspension system, and has a simple mathematical structure, which is based on physically meaningful parameters. As such, it serves as an ideal basis for vehicle identification/parameter estimation.

The dynamic behavior of a vehicle can be characterized most meaningfully by considering the input-output relationship. The input can arise from road roughness or on-board sources or combination of the two. The on-board sources are referred to as those arising from rotating components including the tyre/wheel assemblies, the driveline, and the engine. However, the vehicle properties considered here are those from the road input only. The quarter vehicle model shown in Fig. 4.2.3 includes an unsprung mass representing the axles and associated wheel hardware and a sprung mass representing the vehicle body. The sprung mass is supported on a primary suspension whose stiffness and damping properties are represented by a spring and a damper

respectively. The tyre is represented as a linear spring, although a damper is often included to represent the small amount of damping inherent in the visco-elastic nature of the tyre [119].

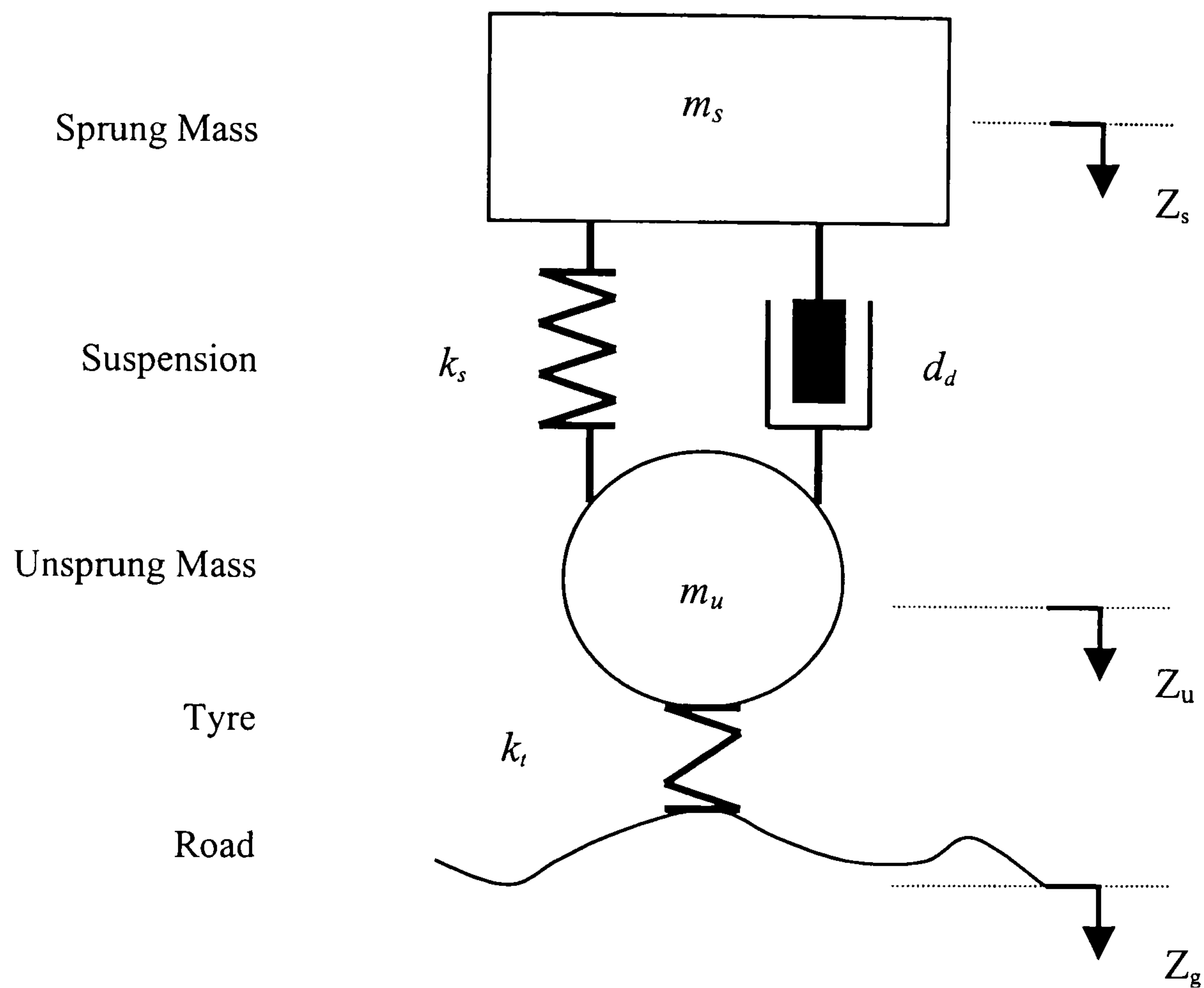


Fig. 4.2.3 A Quarter vehicle model.

The vibrations produced by the quarter vehicle model are as a result of input from the ground input displacement, Z_g , and vertical forces arising from the tyre and suspension components. The motion of the sprung and unsprung masses can be describes by two co-ordinates, Z_s and Z_u , with origins at their static equilibrium positions. The dynamic behavior of the model can then be obtained by applying Newton's second law to the sprung and unsprung masses separately.

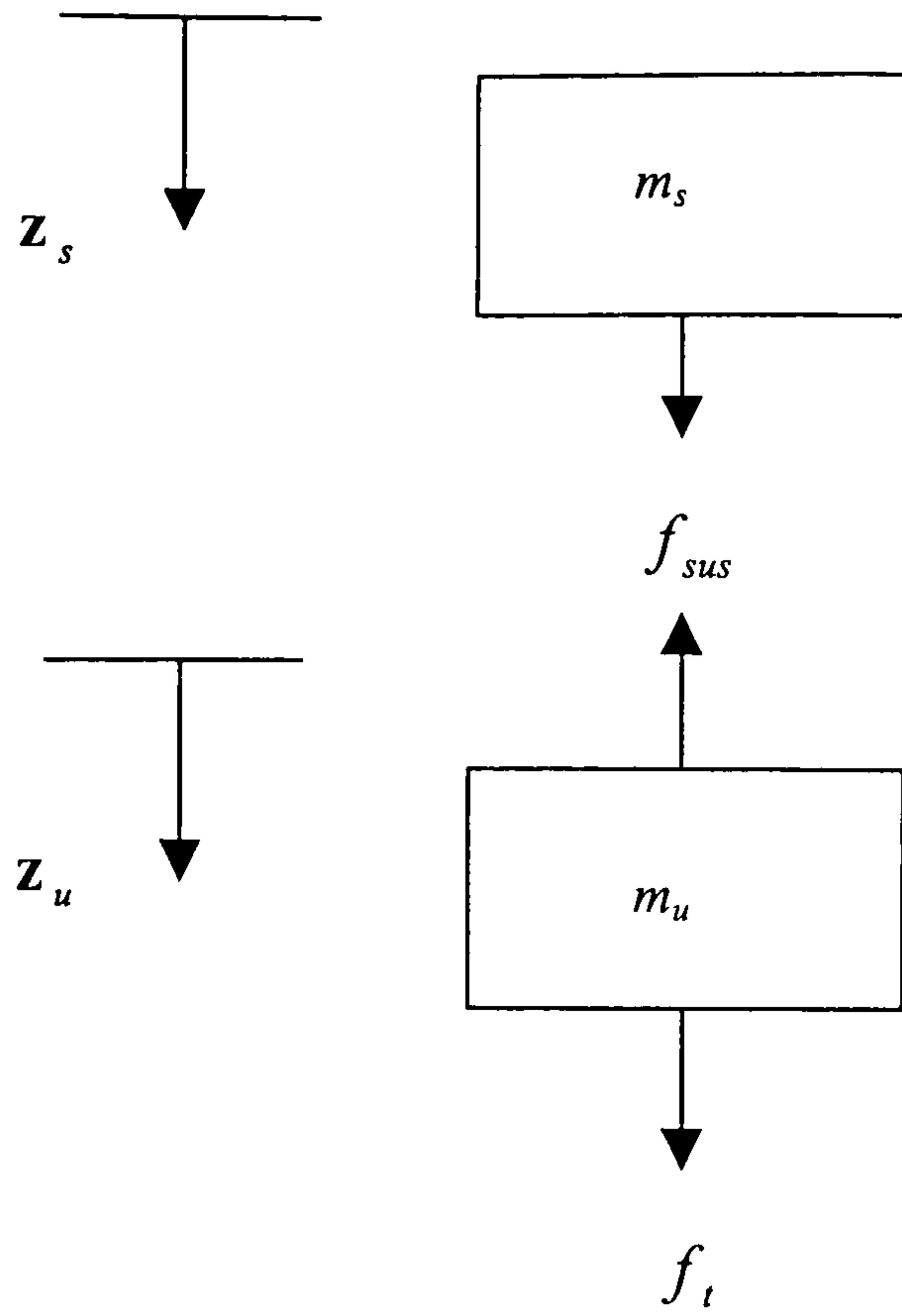


Fig. 4.2.4 A free-body diagram for the quarter vehicle model.

Considering a free-body diagram for each mass in Fig. 4.2.4, the equations of motion of the system are given as follows:

$$m_s \ddot{z}_s = f_{sus} \quad (4.5)$$

$$m_u \ddot{z}_u = -f_{sus} + f_t \quad (4.6)$$

where

$$f_{sus} = f_s + f_d \quad (4.7)$$

$$f_s = -k_s (z_s - z_u) \quad (4.8)$$

$$f_d = -d_d (\dot{z}_s - \dot{z}_u) \quad (4.9)$$

$$f_t = -k_t (z_u - z_g) \quad (4.10)$$

and

z_s = Sprung mass displacement

z_u = Unsprung mass displacement

z_g = Ground input displacement

f_s = Suspension force

f_t = Tyre force

m_s = Sprung mass

m_u = Unsprung mass

k_s = Suspension stiffness

k_t = Tyre stiffness

d_d = Suspension damping coefficient

m_s [Kg]	m_u [Kg]	k_s [kN/m]	k_t [kN/m]	d_d [Ns/m]
200	22	15	230	1500

Table 4.2.1 A list of the representative parameter for the quarter vehicle model

Rearranging to formulate state-space form,

$$\dot{X} = \mathbf{A}X + \mathbf{B}U \quad (4.11)$$

with,

$$X = [Z_s \quad Z_u \quad \dot{Z}_s \quad \dot{Z}_u]^T \quad (4.12)$$

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k_s}{m_s} & \frac{k_s}{m_s} & -\frac{d_d}{m_s} & \frac{d_d}{m_s} \\ \frac{k_s}{m_u} & -\frac{(k_s + k_t)}{m_u} & \frac{d_d}{m_u} & -\frac{d_d}{m_s} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{k_t}{m_u} \end{bmatrix}$$

$$U = Z_g$$

thus the system has two degree of freedom describing vertical motion in terms of four state variables, X . A and B are the system and input matrices respectively. The system is described by two modes, commonly referred to as *body* and *wheel-hop* modes. The wheel-hop frequency is approximately an order higher than the sprung mass resonance frequency (the bounce mode). Practical limits of stroke that can be accommodated within a given vehicle size and suspension envelope constrain the sprung mass natural frequency for most cars to be in the range of 1-1.5 Hz. Performance cars, in which ride is sacrificed for the benefits of handling have a stiff suspension that have the sprung mass natural frequency up to 2-2.5 Hz [99]. A typical unsprung mass resonance frequency (wheel-hop mode) is approximately 10-15 Hz.

The behavior of the quarter vehicle model can be characterized by its poles, that is the eigenvalues of the system matrix A . By taking the linear mass, stiffness and damping constants listed in Table 4.2.1, the eigenvalues of A are given by:

$$\lambda_s = -3.44 \pm 7.85i \quad (4.13)$$

$$\lambda_u = -34.40 \pm 97.37i \quad (4.14)$$

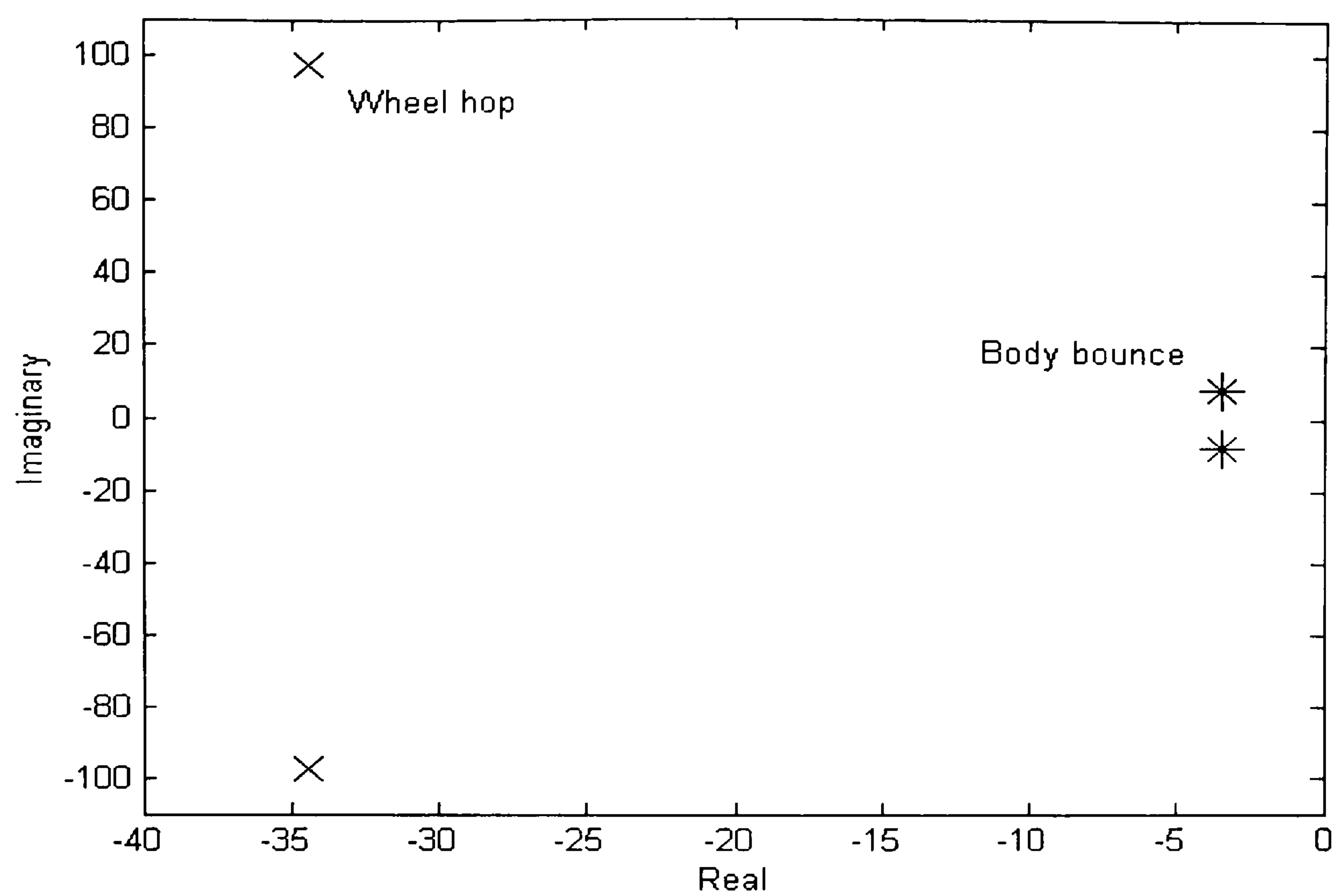
The real and imaginary parts of these eigenvalues are plotted in a so-called s-plane in Fig. 4.2.5(a), where $s = \sigma \pm \varpi_d i$, σ and ϖ_d are the real and imaginary parts respectively. σ and ϖ_d are approximately related to the system damping and natural frequency as follows.

$$\sigma = \zeta \varpi_n \quad \text{and} \quad \varpi_d = \varpi_n \sqrt{1 - \zeta^2}$$

The parameter ζ is called the damping ratio and reflects the level of damping as a fraction of the system critical damping value and, ϖ_d and ϖ_n are called the system damped and undamped natural frequencies respectively.

As can be seen from Fig. 4.2.5(b), the system poles are located at a radius ϖ_n in the s-plane and at an angle of $\theta = \sin^{-1}(\zeta)$. Each pole location in the s-plane determines a particular behaviour of the system response. If σ is negative, the pole is in the left-half plane, the system response decays, and the system is said to be stable. If $\sigma = 0$, the response neither grows nor decays, so stability is a matter of definition. If σ is positive, the response will grow, and the system is said to be unstable. If $\zeta = 0$, the system has no damping, $\theta = 0$ and the damped natural frequency is equal to the undamped natural frequency. The system *rise time*, T_r , which is defined as the time taken for the system to reach the vicinity of its new equilibrium point as a result of step input, may be approximated by $T_r \approx 1.8/\varpi_n$. Similarly, The *settling time*, T_s , which is normally defined as the time taken for the system response to decay so that its magnitude is $\pm 1\%$ of the final value, may be approximated by $T_s \approx 4.6/\sigma$.

(a)



(b)

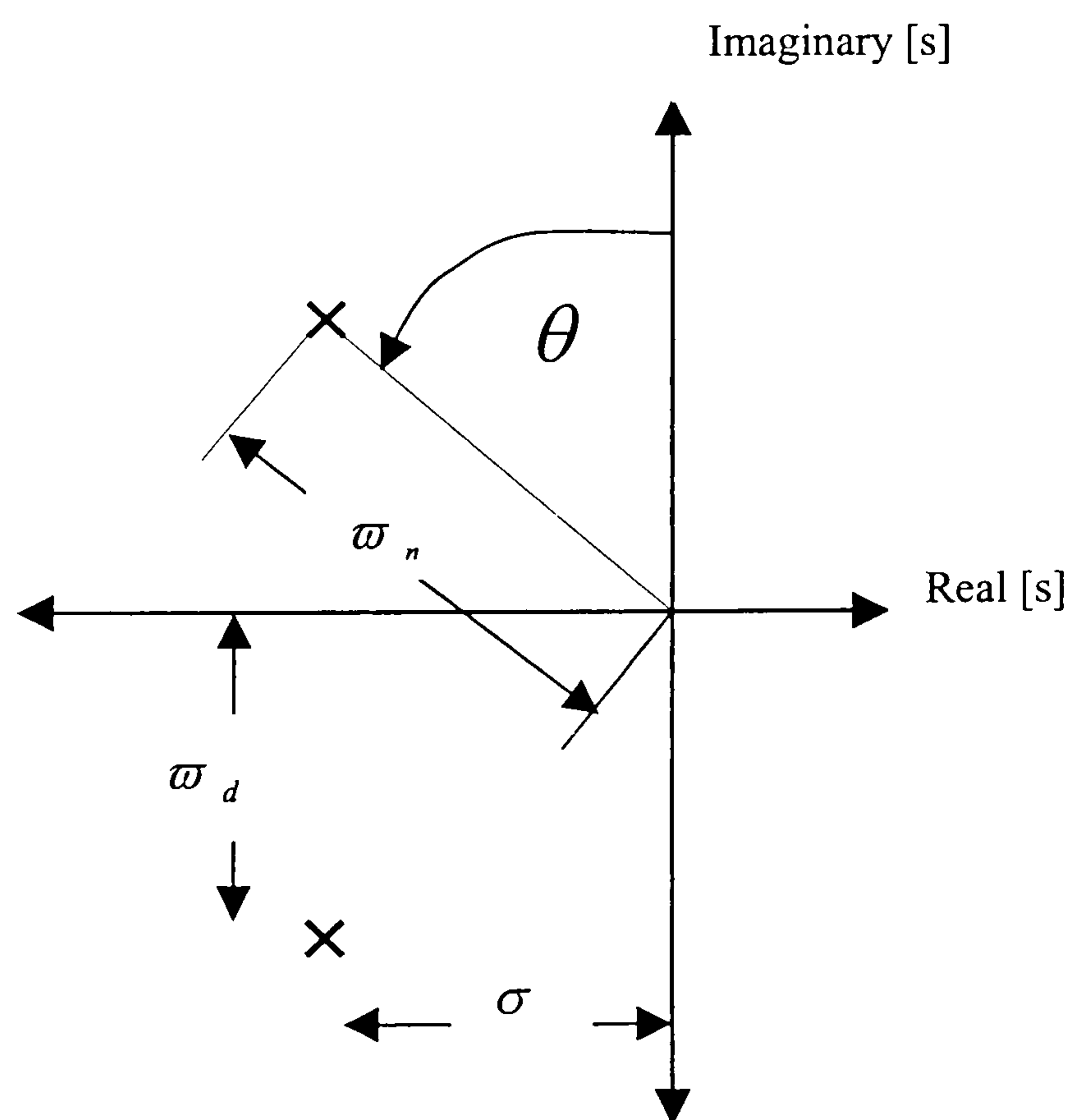


Fig. 4.2.5 The system pole locations (a), and a sketch of a pair of complex poles (b).

From equations (4.10) and (4.11), the undamped natural frequency in Hertz., f_n , the rise time and settling time for each mode are then given by,

	<i>body bounce</i>	<i>wheel-hop</i>
$f_n = \frac{\sqrt{\sigma^2 + \varpi_d^2}}{2\pi}$	1.36 Hz	16.4 Hz
$T_r = \frac{1.8}{\varpi_n}$	0.21 secs	0.017 secs
$T_s = \frac{4.6}{\sigma}$	1.34 secs	0.13 secs

Fig. 4.2.6 illustrates the system modes, showing time histories of the response for the sprung and unsprung masses to an impulsive road displacement input. It also illustrates modal settling times, and shows that the energy from an isolated road disturbance is absorbed initially through oscillations of the unsprung mass. These in turn excite slower and less extreme oscillations of the sprung mass.

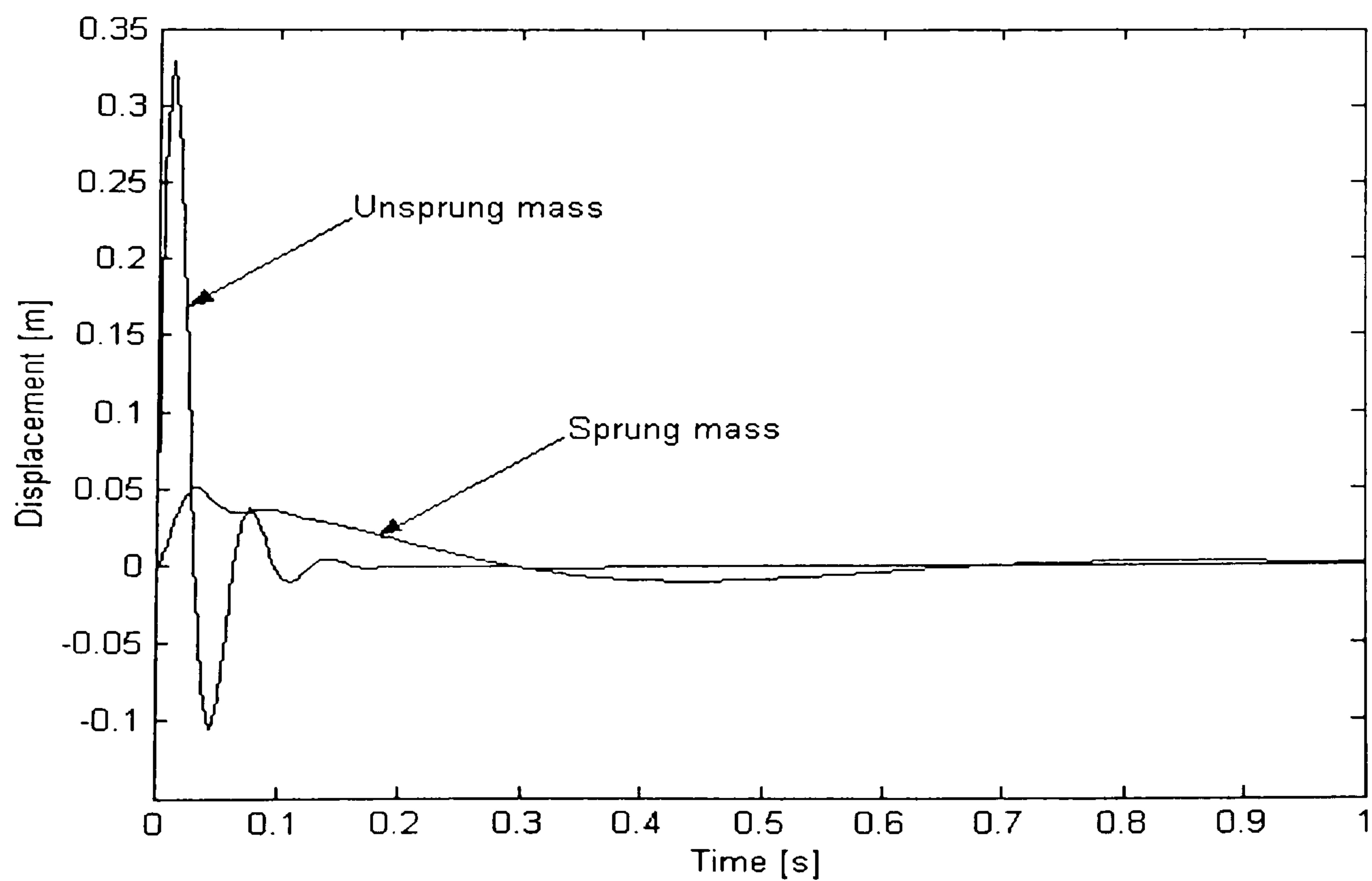


Fig. 4.2.6 A impulse response of the quarter vehicle model.

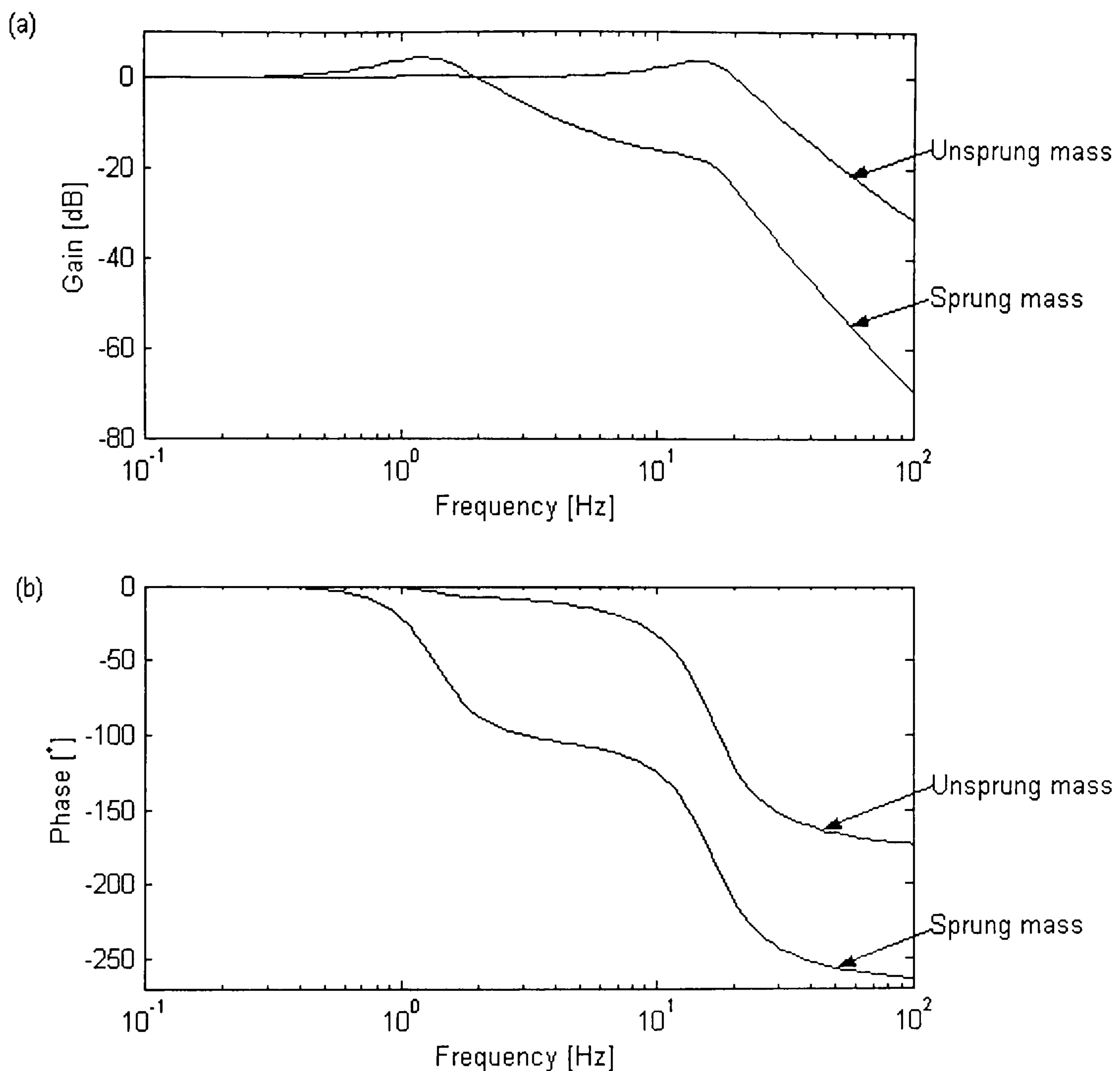


Fig. 4.2.7 A frequency response of the quarter vehicle model.

The system frequency response is shown in Fig. 4.2.7(a) and (b). Fig. 4.2.7(a), these suggest that the sprung mass motion is entirely dominated by the bounce mode, whereas the unsprung motion is excited in both modes. However, the unsprung mass oscillates across a broader bandwidth and its energy is spent principally at the higher frequencies. It is evident from Fig. 4.2.7(b) that at low frequencies, where phase difference between the two curves is very small the sprung and unsprung masses move as an integral unit. As the frequency of excitation increases, the phase

difference also increases and approaches the maximum theoretical value of 90° at about a frequency of 4 Hz.

4.2.2 Influence of vehicle parameter variation to its response

To study the influence of variation in the vehicle parameters, simulations were carried out by varying each one individually, while holding the others constant at the values given in Table 4.2.1. The parameters were varied as follows:

m_s between 100 - 300 kg,	m_u between 10 - 50 kg,
k_s between 5 - 30 kN/m,	k_t between 100 - 500 kN/m,
d_d between 500 - 3000 Ns/m.	

An indication of the effect of parameter variation was obtained by observing the system frequency responses as well as the system pole locations, as presented in Fig. 4.2.8 – Fig. 4.2.12. Part (a) of the figures show Bode plots of the magnitude and phase of the transfer function between the sprung mass and ground input for different parameter values. Similarly, the magnitude and phase of the transfer function between the unsprung mass and ground input are shown in part (b). The system poles, which are usually expressed in complex conjugate pairs, are plotted in part (c). However, for clarity, only poles with a positive imaginary value are shown. The poles at the bounce mode are represented by ' * ' signs and ' x ' signs representing poles at the wheel-hop mode. The arrow signs indicate the direction of increase of each parameter of interest.

From Fig. 4.2.8 (a) and (b), as the sprung mass increases, the resonance at the bounce mode occurs at a lower frequency, and at a slightly higher frequency for the wheel-hop mode. With a heavier sprung mass, the gain of both the sprung and unsprung masses at frequencies beyond the sprung mass resonant point are smaller when compared with that of a lighter sprung mass but are higher at frequencies below the sprung mass resonant point. This suggests that the transmission from the ground input to the sprung and unsprung masses is high at low frequencies and low at high frequencies.

It is also seen from Fig. 4.2.8(c) that, as the sprung mass increases, the poles at the bounce mode move to the right hand side of the figure. The poles at the wheel-hop mode move upwards and are close to each other. These, again, suggest that the bounce mode takes place at a lower frequency and the wheel-hop mode at a slightly higher frequency, recalling from Fig. 4.2.5 that the undamped natural frequency may be determined by the distance from the origin to the pole coordinates. Again from Fig. 4.2.5, it can also be said that a heavier sprung mass reduces the system damping particularly at the bounce mode. The rise time of the sprung mass, which is approximately inversely proportional to the undamped natural frequency, increases since the natural frequency decreases. In other words, the response of the sprung mass to the input is slower as the sprung mass is made heavier. This is also in a agreement with reduction of the phases observed in Fig 4.2.8 (a) and (b), with greater variation seen on the phase of the sprung mass.

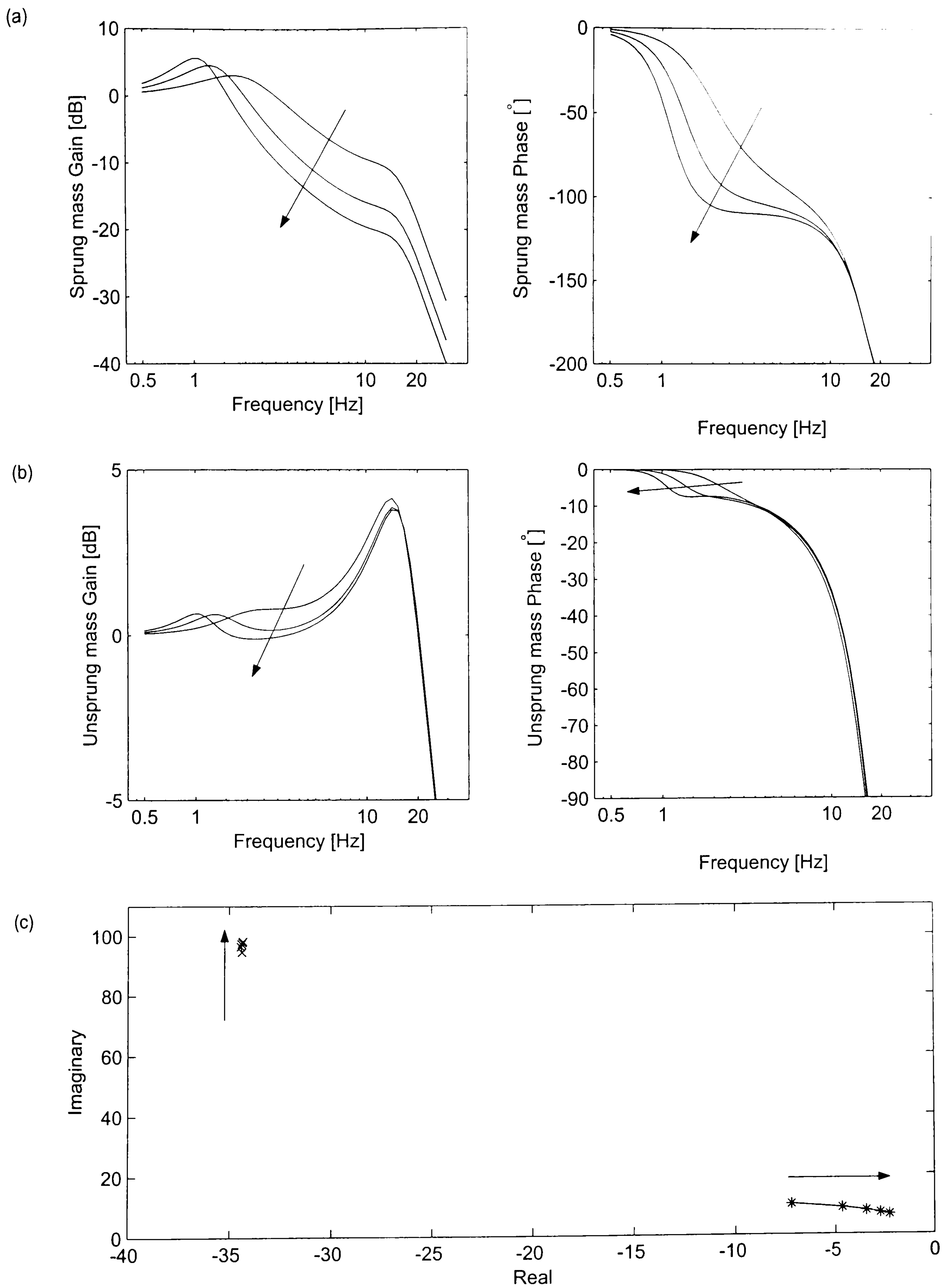


Fig. 4.2.8 Influence of sprung mass variation; (a) on sprung mass response, (b) on unsprung mass response, (c) system pole location.

However, change in unsprung mass has very little influence on the body bounce mode. This is clearly seen from Fig. 4.2.9 (a), (b) and (c). The system poles at the bounce mode are very close together and are seen as one mark at the lower right corner of the plot. Increasing the unsprung mass causes the wheel-hop resonance to take place at a lower frequency, with reduction of the gain observed at frequencies beyond the wheel-hop mode, and higher gain observed between the two modes. A heavier unsprung mass causes the response to slow down and also reduces the damping. This is indicated by Fig. 4.2.9 (c) as the pole locations moving downwards and to the right hand side of the figure.

The suspension spring has a significant influence on both gain and phase of the sprung and unsprung mass responses at frequencies below the wheel-hop resonance. Increasing the suspension spring stiffness causes the two resonance modes to occur at slightly higher frequencies. As far as the bounce mode is concern, a stiffer suspension spring generates greater forces resulting in higher sprung and unsprung mass gains. At frequencies between the two modes, increase in suspension spring stiffness results in increase in the sprung mass gain, and a reduction in that of the unsprung mass, as seen from Fig. 4.2.10(a) and (b). It is observed from Fig. 4.2.10(c) that, as the suspension spring stiffness increases, poles at the bounce mode move upwards and to the right such that the poles are almost the same distance from the origin point. This means that the damping of the system at the bounce mode increases, and the effect of suspension spring variation at the bounce mode is greater on the damping, rather than the natural frequency. While poles at the wheel-hop move upwards and to the left, indicating an increase in the unsprung mass natural frequency, and an increase in the damping,

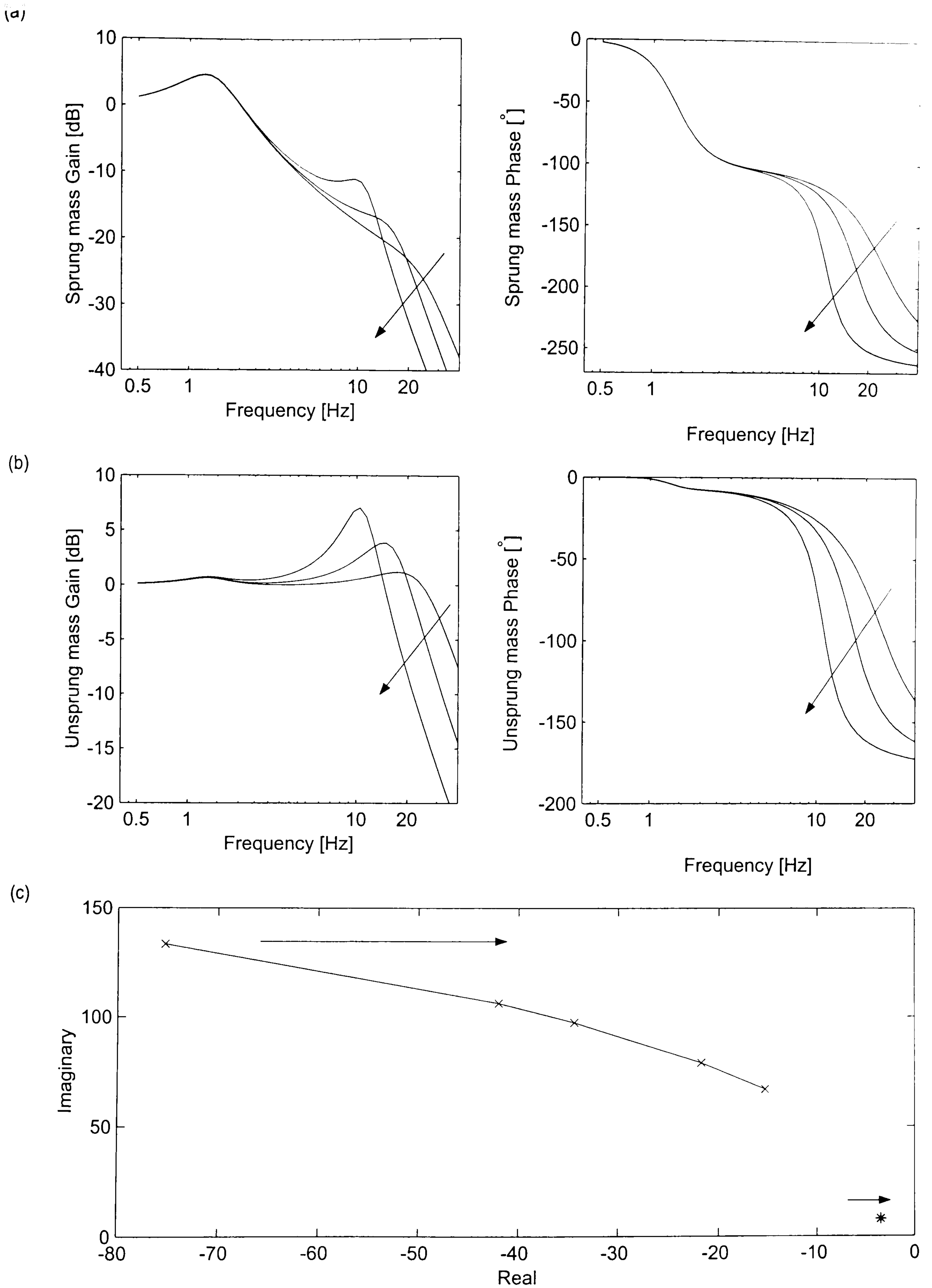


Fig. 4.2.9 Influence of unsprung mass variation; (a) on sprung mass response, (b) on unsprung mass response, (c) system pole location.

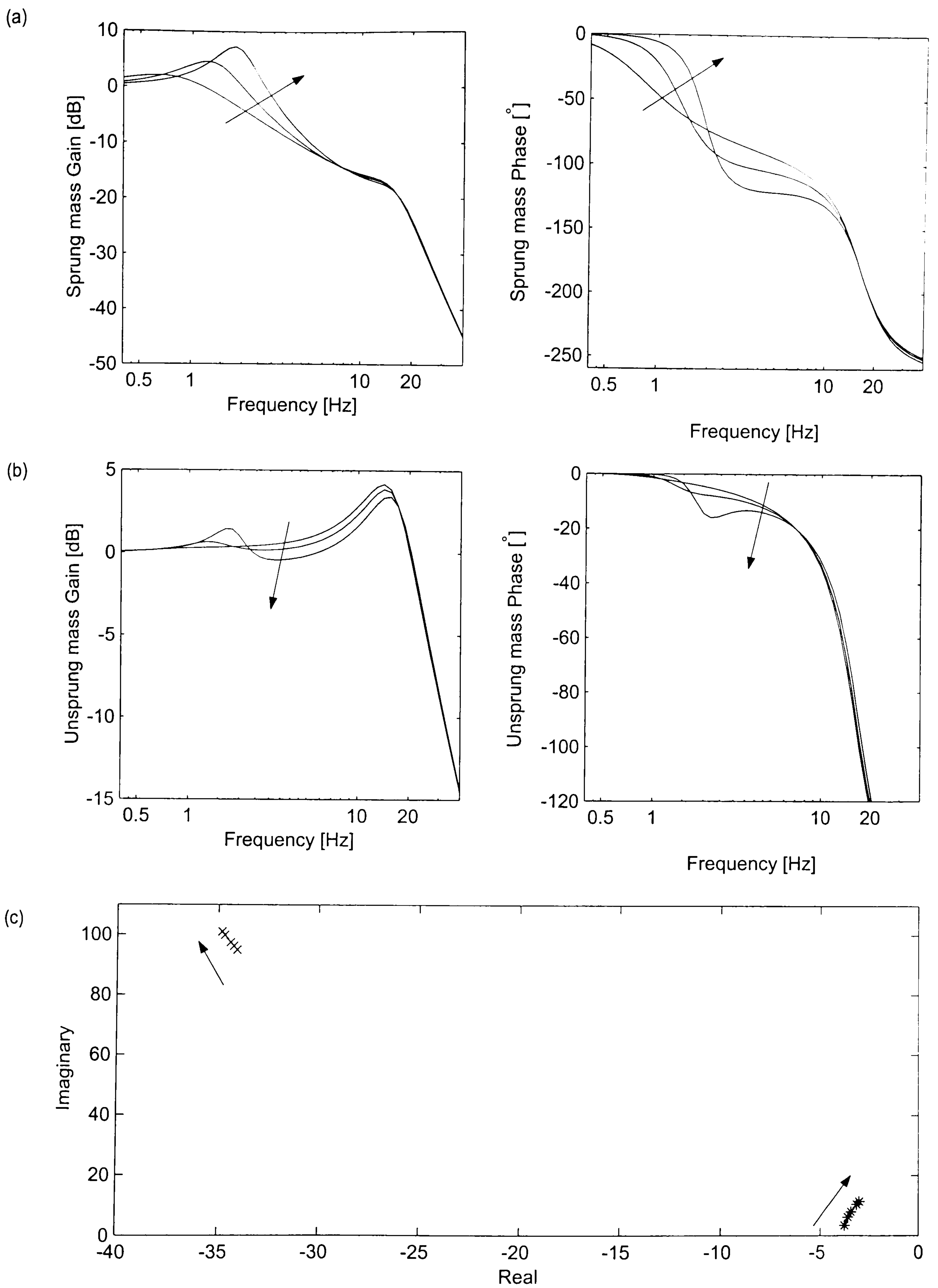


Fig. 4.2.10 Influence of spring stiffness variation; (a) on sprung mass response, (b) on unsprung mass response, (c) system pole location.

however by only a small amount. It then implies that change of suspension spring stiffness tends to influence the wheel-hop resonance frequency rather more than the damping.

It is seen from Fig. 4.2.11(a) and (b) that increase in tyre stiffness produces higher sprung and unsprung mass gains and phases at frequencies beyond wheel-hop resonance frequency. The resonance frequency at the wheel-hop mode increases, as the tyre stiffness is greater. From Fig. 4.2.11(c), pole locations at the wheel-hop mode move upward indicating a higher natural frequency and damping ratio at the mode. Poles at the bounce mode are close together, thus it can be concluded that variation of tyre stiffness has small influence on the bounce mode, and a significant influence to the wheel-hop mode.

It is observed from Fig. 4.2.12(a) and (b) that the damping ratio variation significantly affects the magnitudes at both resonance frequencies, and thus generates a great phase shift. Furthermore, the two natural frequencies are almost unchanged. In Fig. 4.2.12(c), as the damping ratio increases the system poles move upwards and to the right hand side of the figure suggesting an increase in the damping as expected.

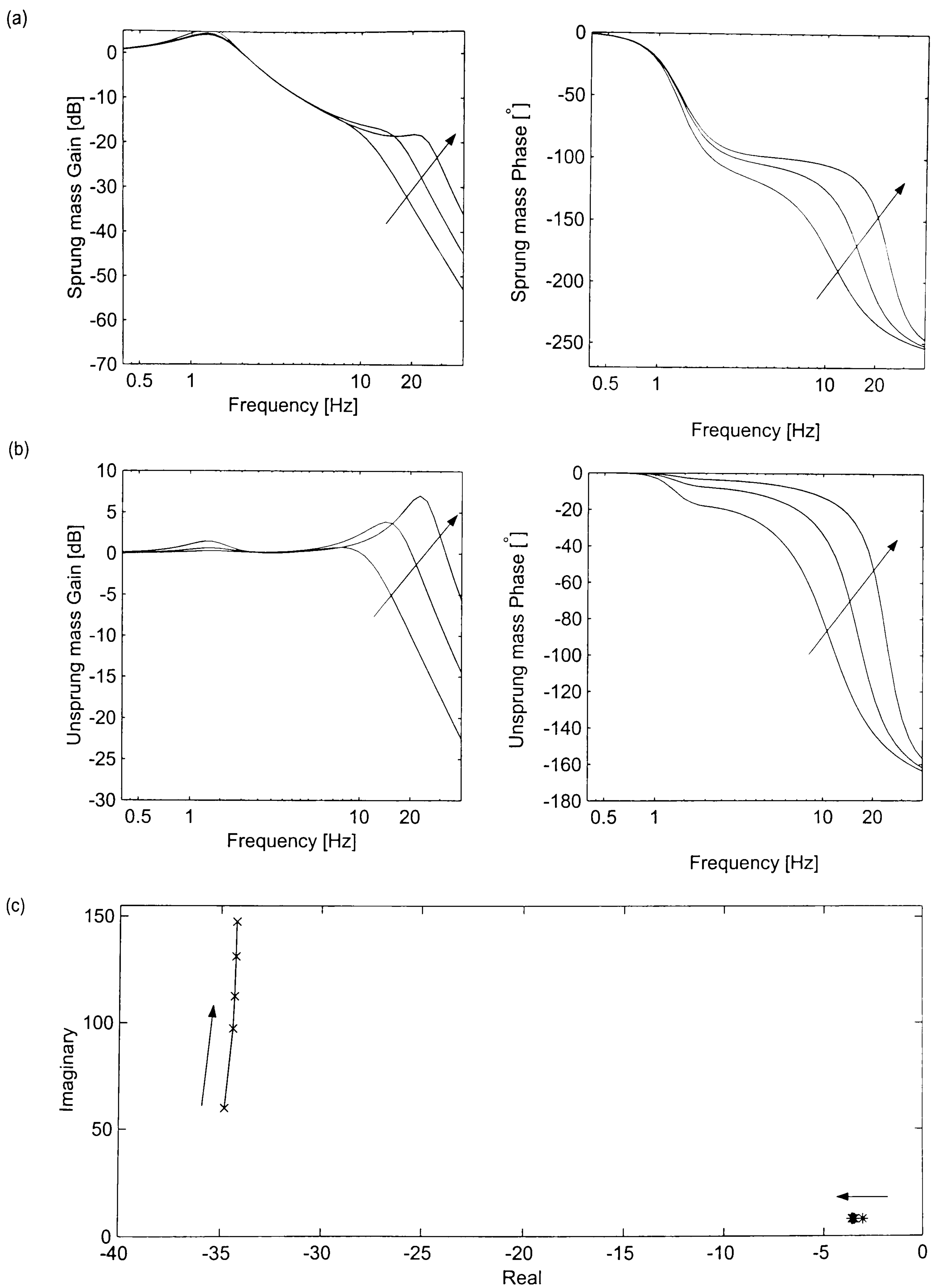


Fig. 4.2.11 Influence of tyre stiffness variation; (a) on sprung mass response, (b) on unsprung mass response, (c) system pole location.

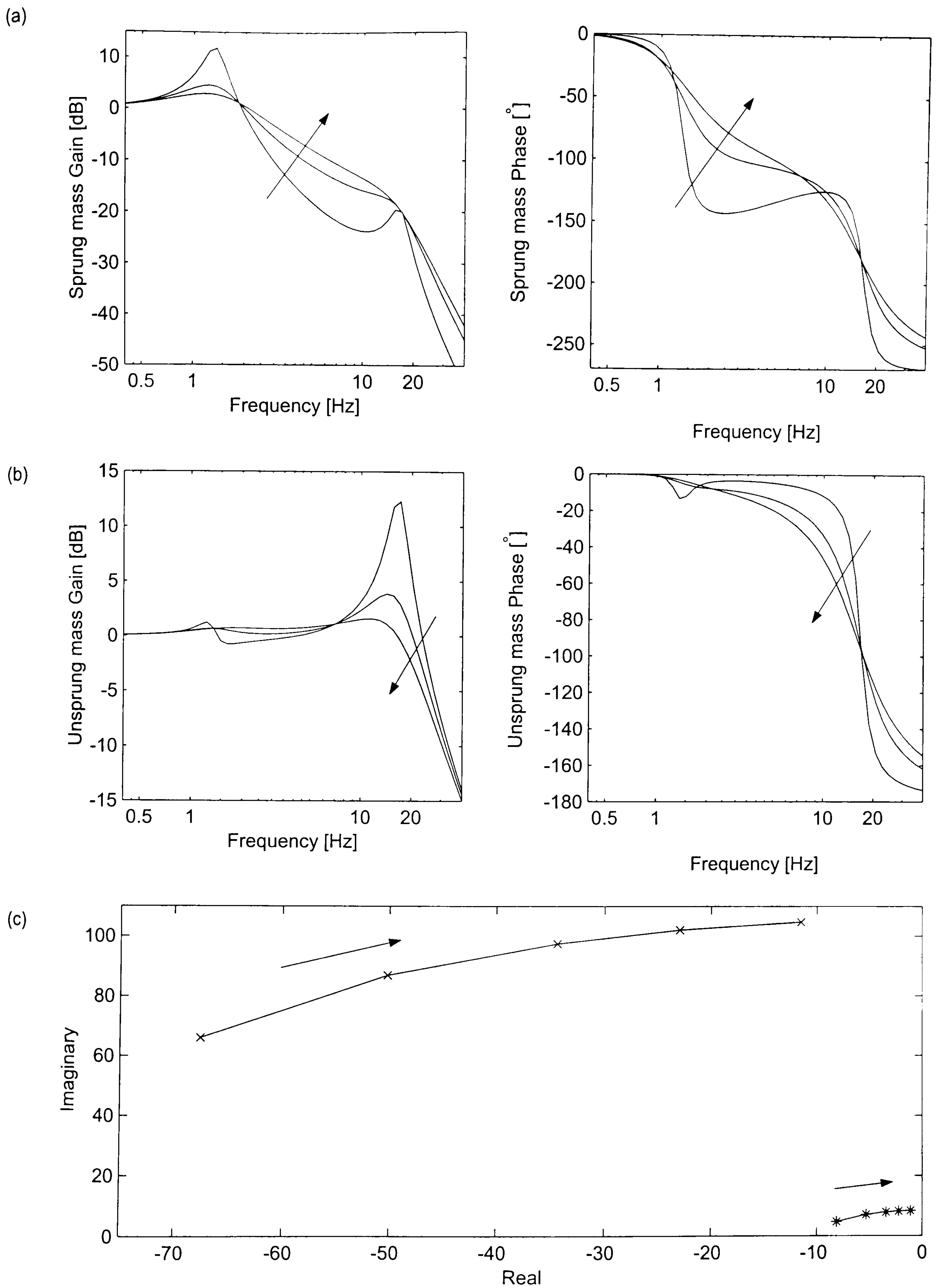


Fig. 4.2.12 Influence of damping rate variation; (a) on sprung mass response, (b) on unsprung mass response, (c) system pole location.

To summarise, the simulation results show that each parameter contributes to a certain amount of change, in a particular direction, in the response of the vehicle. Variation of the sprung mass has more influence on the bounce mode, whereas changes in the unsprung mass and tyre stiffness variation are more visible at the wheel-hop mode. Suspension spring stiffness has an influence on both modes. Variation of damping rate has a larger influence on the resonance frequencies than the other parameters.

4.2.3 Estimation of model transfer function.

In the last section the response was primarily investigated in the frequency domain using Bode plots, this is mainly used for Linear Time Invariant (LTI) systems. In this section, a method of estimating the model transfer function for a nonlinear system is developed. The method is validated and its use and limitations for the quarter vehicle model are studied and discussed.

First, consider a stable single input-single output (SISO), LTI system with output $y(t)$, input $u(t)$ and system transfer function G . The convolution integral is

$$y(t) = \int_0^t G(t - \tau) u(\tau) d\tau \quad (4.15)$$

where τ is an auxiliary time variable. If the input is a complex vector at the frequency ω radian per second, $u(\tau) = A_u e^{j\omega\tau}$, and defining $\zeta = t - \tau$ then

$$\begin{aligned}
y(t) &= \int_0^t G(t-\tau) A_u e^{j\omega\tau} d\tau \\
&= \int_0^t G(\varsigma) A_u e^{j\omega(t-\varsigma)} d\tau \\
&= A_u e^{j\omega t} \int_0^t G(\varsigma) e^{-j\omega\varsigma} d\varsigma
\end{aligned}$$

The integral can be represented as the sum of the two infinite integrals, using the definition of the Laplace transform

$$\begin{aligned}
y(t) &= A_u e^{j\omega t} \left[\int_0^\infty G(\varsigma) e^{-j\omega\varsigma} d\varsigma - \int_t^\infty G(\varsigma) e^{-j\omega\varsigma} d\varsigma \right] \\
&= A_u e^{j\omega t} G(s) \big|_{s=j\omega} - A_u e^{j\omega t} \int_t^\infty G(\varsigma) e^{-j\omega\varsigma} d\varsigma \\
&= y_{ss}(t) + y_t(t)
\end{aligned} \tag{4.16}$$

Thus $y(t)$ is the sum of two terms; the transient response of the system, $y_t(t)$ and the steady state response, $y_{ss}(t)$. For any $t \geq 0$ the magnitude of $y_t(t)$ is

$$\begin{aligned}
|y_t(t)| &= \left| A_u e^{j\omega t} \int_t^\infty G(\varsigma) e^{-j\omega\varsigma} d\varsigma \right| \\
&= A_u |e^{j\omega t}| \cdot \left| \int_t^\infty G(\varsigma) e^{-j\omega\varsigma} d\varsigma \right| \\
&= A_u \cdot \left| \int_t^\infty G(\varsigma) e^{-j\omega\varsigma} d\varsigma \right| \\
&\leq A_u \cdot \int_t^\infty |G(\varsigma) e^{-j\omega\varsigma}| d\varsigma \\
&= A_u \cdot \int_t^\infty |G(\varsigma)| d\varsigma
\end{aligned}$$

If a bounded input-bounded output (BIBO) convolution system is assumed, then $y_t(t)$ has a finite value. As $t \rightarrow \infty$ the value of the integral become small, $\lim_{t \rightarrow \infty} (y_t(t)) = 0$, so

that finally the output of the system is

$$y(t) = A_u e^{j\omega t} G(s)|_{s=j\omega} + y_t(t)$$

and the steady-state behaviour of the system, as $t \rightarrow \infty$, is

$$y_{ss}(t) = A_u e^{j\omega t} G(j\omega) \quad (4.17)$$

That is, the steady-state response is the transfer function evaluated at $j\omega$ multiplied by the sinusoidal input. Thus, from the preceding derivation, the steady state response of the system excited by a sinusoid provides the transfer function of the system. To give an example, the simplest case, consider a simple periodic signal, $y(t)$, generated by a stable LTI system excited by an input signal $u(t)$ comprising of a single, known frequency ω with a known amplitude of A_u , where

$$u(t) = A_u \cos(\omega t) \quad (4.18)$$

$$y(t) = A_{ym} + A_{yc} \cos(\omega t) + A_{ys} \sin(\omega t) \quad (4.19)$$

That is, $y(t)$ comprised of three components; a mean level A_{ym} depending on system initial condition and transient characteristic, and the cosine and sine components.

Given that $\omega = \frac{2\pi}{T}$, where T is the period in second, and using trigonometric

identities, an alternative form of $y(t)$ can be written in cosine component and a phase shift as follow

$$y(t) = A_{ym} + A_y \cos(\omega t - \phi) \quad (4.20)$$

where

$$A_y = \sqrt{A_{yc}^2 + A_{ys}^2} \quad (4.21)$$

$$\phi = \tan^{-1} \left(\frac{A_{ys}}{A_{yc}} \right) \quad (4.22)$$

The coefficients A_{ym} can be estimated in minimum square error sense as;

$$A_{ym} = \frac{1}{T} \int_{t-T}^t y(t) . d(t)$$

By changing the integral limit and replacing $y(t)$ by $y(\omega t)$ and t by ωt , the expression for A_{ym} becomes

$$A_{ym} = \frac{1}{2\pi} \int_{-\pi}^{\pi} y(\omega t) . d(\omega t)$$

Similarly, A_{yc} , and A_{ys} can be estimated as follows;

$$A_{yc} = \frac{2}{2\pi} \int_{-\pi}^{\pi} y(\omega t) \cos(\omega t) . d(\omega t)$$

$$A_{ys} = \frac{2}{2\pi} \int_{-\pi}^{\pi} y(\omega t) \sin(\omega t) . d(\omega t)$$

When $y(\omega t)$ is sampled, the integration reduces to summation. If sampled at a constant rate of F_s , then the number of sampled data points in one period, n_p , is equal to F_s . The coefficients are then rewritten as follows

$$A_{ym} = \frac{1}{n_p} \sum_{k=1}^{n_p} y(\omega t_k) \quad (4.23)$$

$$A_{yc} = \frac{2}{n_p} \sum_{k=1}^{n_p} y(\omega t_k) \cos(\omega t_k) \quad (4.24)$$

$$A_{ys} = \frac{2}{n_p} \sum_{k=1}^{n_p} y(\omega t_k) \sin(\omega t_k) \quad (4.25)$$

where $t_k = \frac{k}{F_s}$, $k = 1, 2, 3, \dots, n_p$. Then the single cosine component, A_y , together with the phase shift, ϕ , in equations (4.21)-(4.22) are determined. Since assumed priori known input magnitude and phase, for one particular frequency of interest ω , using correlation method the system transfer function is therefore estimated as

$$\text{Gain} = \frac{A_y}{A_u} \quad (4.26)$$

$$\text{Phase lag} = \phi. \quad (4.27)$$

The development just derived above can be used to carry out a number of frequency points, so that the frequency response can be determined. Next, a simulation study of the method described above is validated against the results from the standard linear frequency response methods. The simulation study is carried out on the quarter

vehicle model whose parameters are given in Table 4.2.1. Plots of gain and phase obtained from the two methods are shown on the same graphs in Fig. 4.2.13. The differences between them are unnoticeable small. The gain and phase errors are plotted against frequency in Fig. 4.2.14. The method gives a very good result at low frequencies and up to approximately 10 Hz, after which larger result. This is justified from the fact the method employs a simple strategy with a constant sampling rate. For the same number of cycles, thus number of data points for one particular frequency decreases as the frequency increases. The errors then arise from the summation when calculating the coefficients A_{yc} and A_{ys} . The choice of sampling rate is a compromise between the estimation accuracy and total time required when obtained the model outputs. A higher sampling rate would reduce the errors, with an expense of more time needed. A varying sampling rate may be employed with increased complexity with the problem that additional time is required.

For the method developed, the errors are within ± 0.5 dB for the sprung and unsprung mass gain, and approximately $\pm 0.5^\circ$ for the phase errors. Though the errors increase with frequency, it can be said that, in the frequency range of interest, the errors are acceptable with respect to the values of gain and phase. Hence, the method is justified bearing in mind its simplicity.

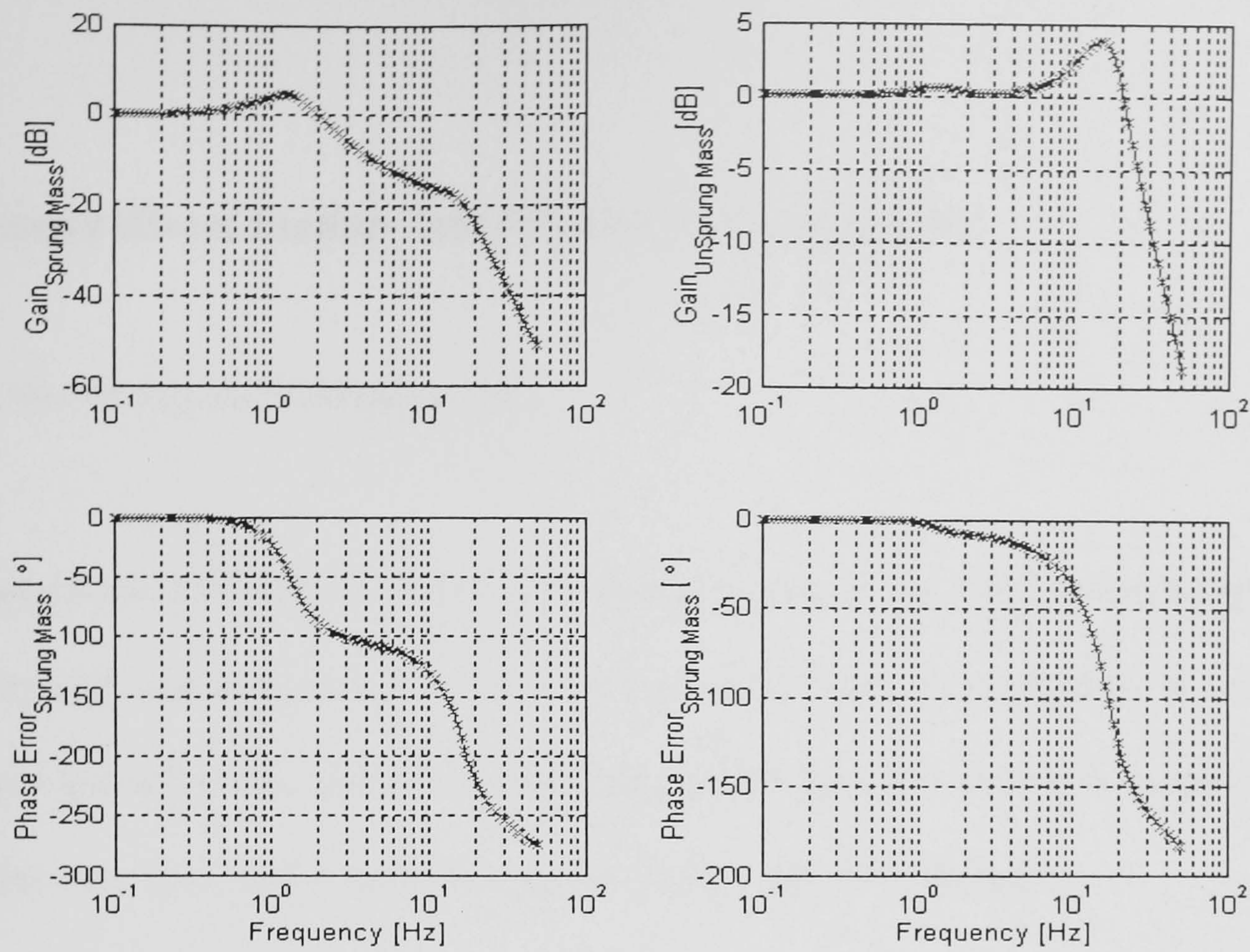


Fig. 4.2.13. Validation of the transfer function estimation method against Bode method.

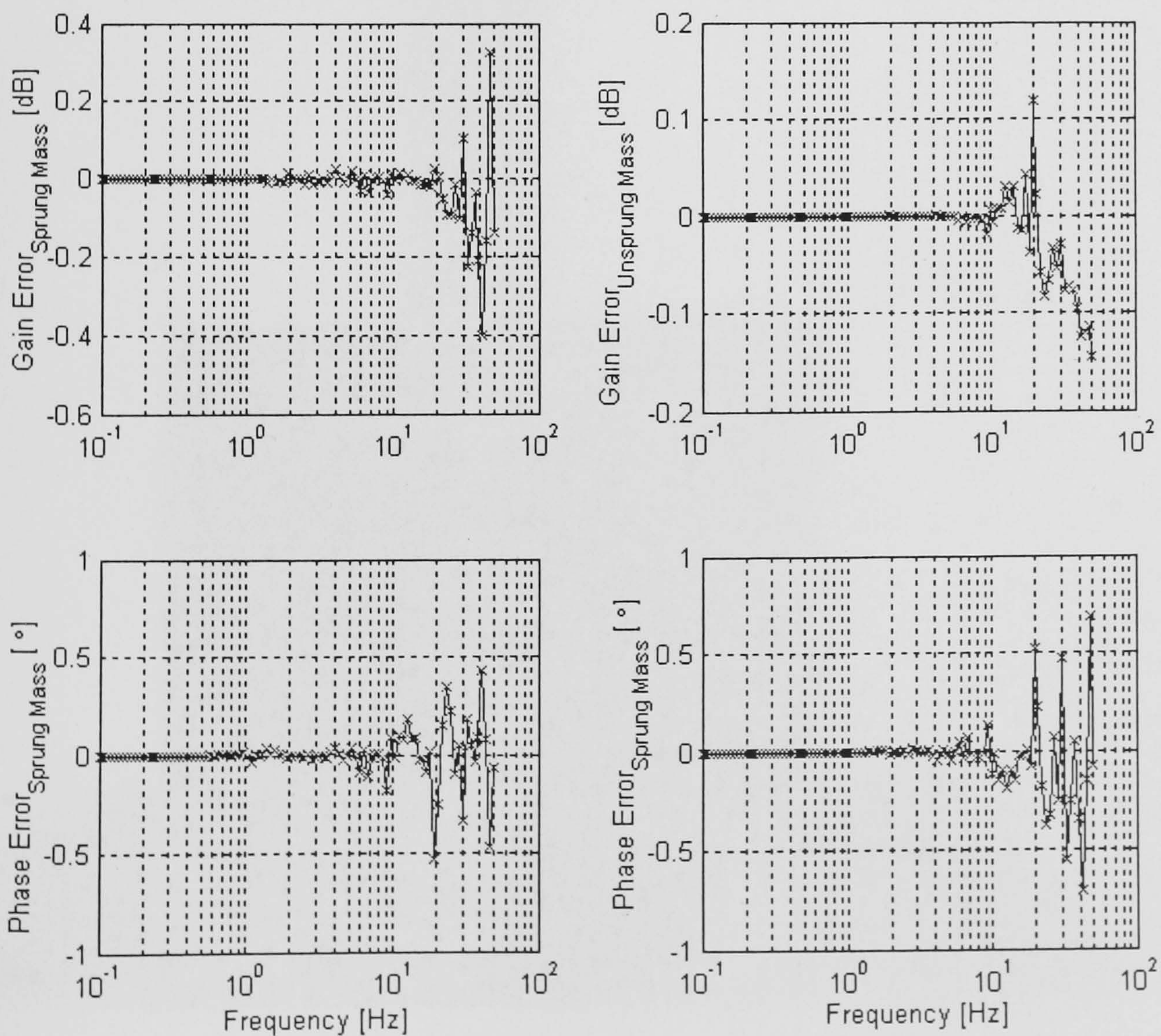


Fig. 4.2.14. Validation of the transfer function estimation method against Bode method; Gain and Phase errors

4.3 Single Wheel Station Test Rig and Instrumentation

4.3.1 The test rig and Instrumentation

A photograph of the Single Wheel Station rig is shown in Fig. 4.3.1, and a diagram of the test rig and instrumentation is given in Fig. 4.3.2. The rig comprises of a hydraulic actuator located at the bottom to simulate ground inputs, a load cell, which measures dynamic tyre force that is directly mounted on the top of the actuator.

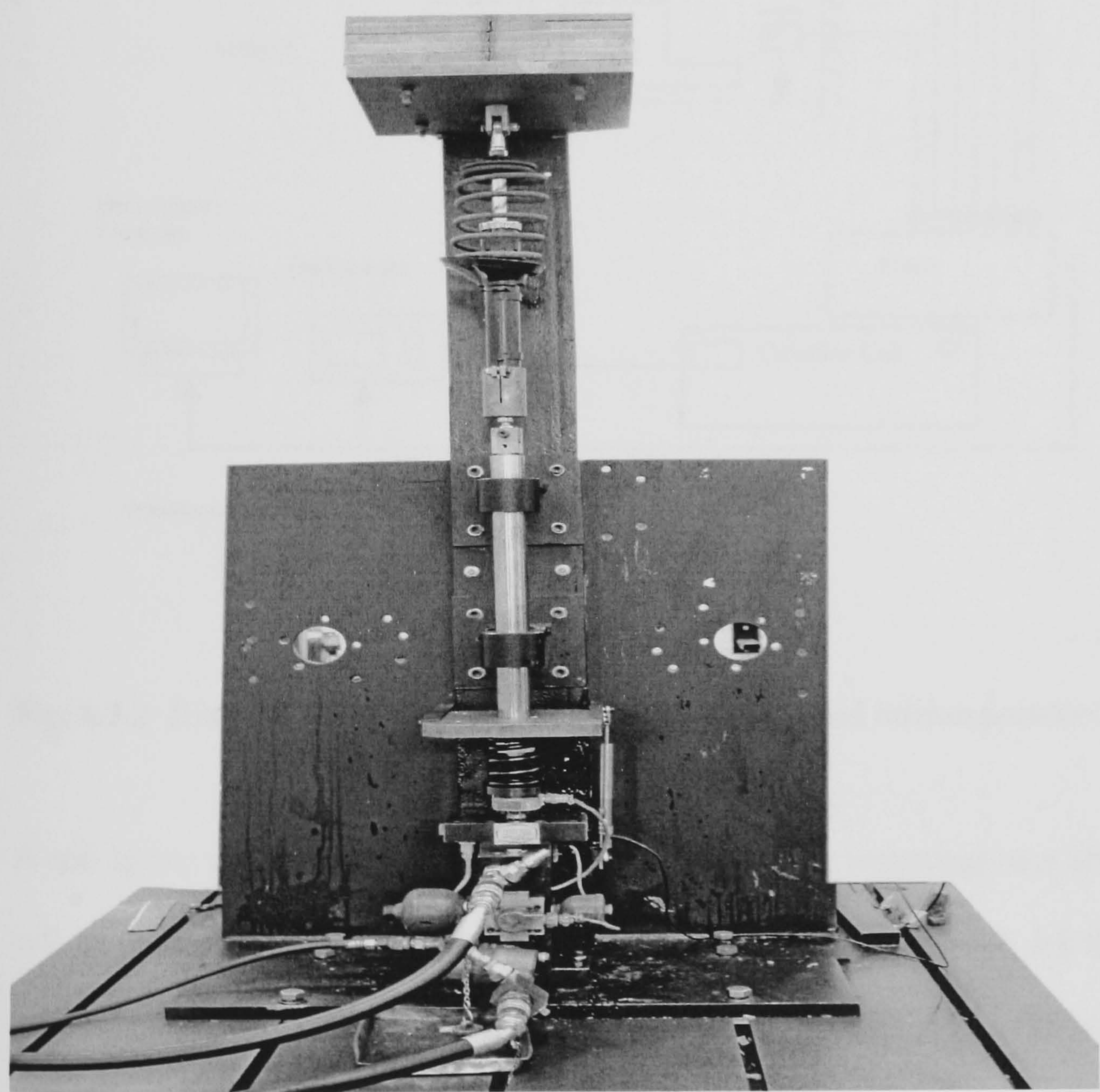


Fig. 4.3.1 Photo of the Single Wheel Station.

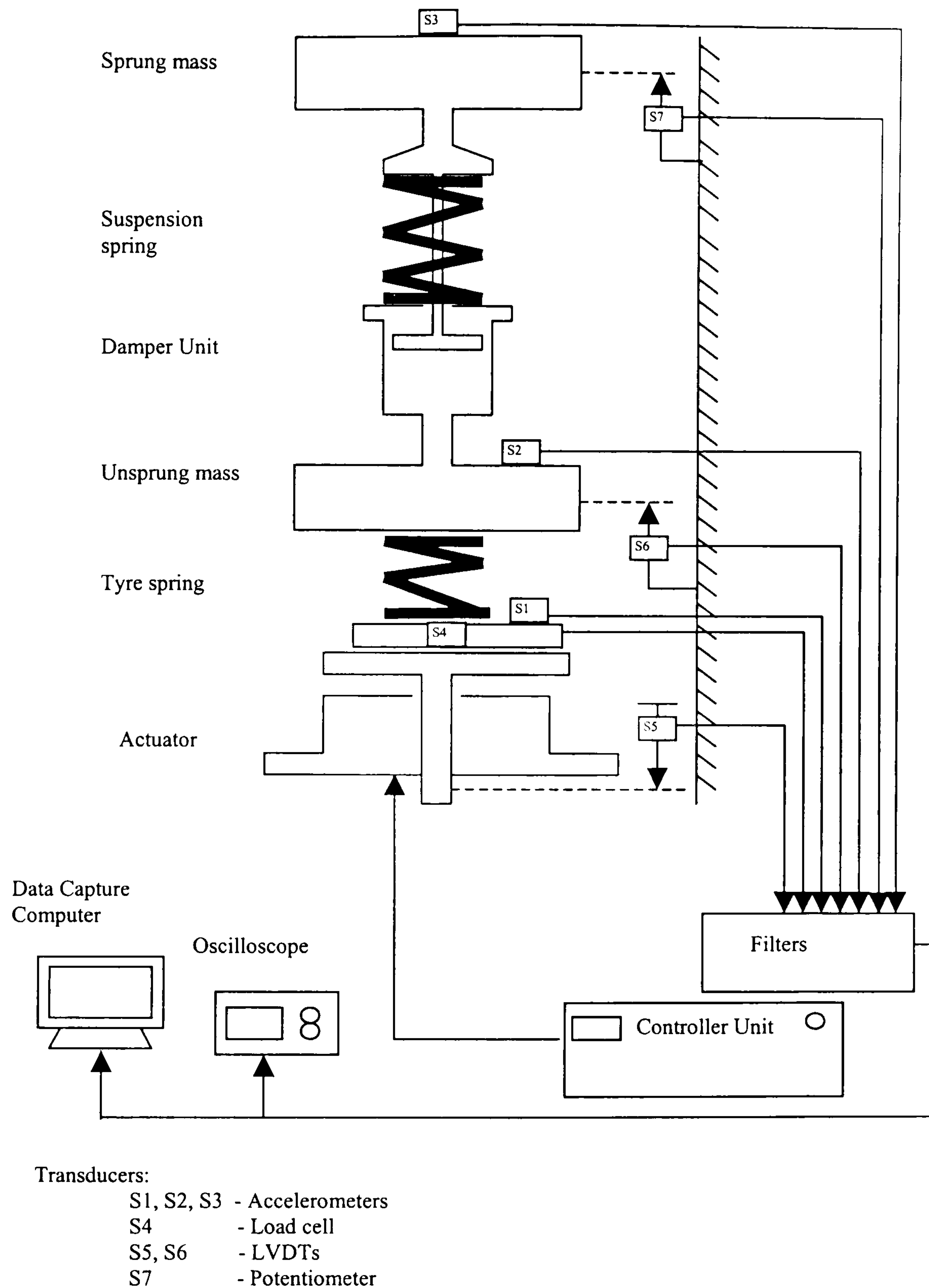


Fig. 4.3.2 Diagram of the Single Wheel Station test rig and instrumentation.

A coil spring representing the tyre is located between the unsprung mass and the load cell. It should be born in mind that, for a full vehicle, it is difficult for sprung and unsprung masses to be clearly distinguished. In this section, the unsprung mass is referred to the mass above the tyre spring and up to and including the bottom part of the damper unit, and the sprung mass includes the piston of the damper and those

items above it. The damper is mounted between the unsprung and sprung masses and provides damping as well as location for the suspension spring.

The movements of the sprung and unsprung masses are guided vertically by the rods, which are supported in bearings. To prevent the sprung and unsprung masses from rotating when excited, a pair of guide rollers is attached to the masses. The actuator has a dynamic stroke of ± 12.5 mm, and the movement of the actuator is controlled by a digital servo control system. The controller, Kelsey 7500, is capable of generating the input as well as passing the input signal from an arbitrary source to the rig. The input displacement of the actuator was measured using a linear voltage displacement transducer (LVDT) attached to the bottom end of the actuator. Another LVDT was used to measure displacement of the unsprung mass. The displacement of the sprung mass was measured by using a potentiometer. Three accelerometers were attached to the sprung mass, unsprung mass and the load cell to measure the accelerations of the sprung and unsprung masses, and ground input respectively.

The input signal chosen for the test was a swept sine, with its magnitude set to maximum the allowable motion of the actuator. The frequencies of interest of the system are below 30 Hz for this study. All of the response signals from the system were pre-filtered at 30 Hz by a 4th order low pass filter to minimize high frequencies noise such as those which arose from electrical noise and the hydraulic pump. A constant sampling rate of 500 Hz was founded to be sufficient and reasonable compromise between the signal reconstruction and time taken to run the estimation algorithm.

4.3.2 Instrument Calibrations

Measurement errors have unwanted effect on the parameter to be identified, and thus the calibration of test instrumentation was carried out carefully.

To calibrate the LVDT, one of its ends was held fixed, and the other was moved to different distances. The readings of distance against voltage were recorded and used to obtain a look-up table for each LVDT. The calibrations were also checked that they agreed with the specification given by the manufacture.

The load cell was calibrated using a hydraulic press and proofing ring. It was loaded with a series of known loads, and the voltage readings were recorded. A look-up table of load and voltage for the load cell was then formulated and used to obtain the dynamic tyre force.

The calibration of the accelerators was performed into two stages. First, the dynamic calibration was performed, by placing all the accelerometers on one of the hydraulic actuator of the Four Poster Rig, next to the calibrated accelerometer. They were then excited by a swept sine signal through a frequency range and at each frequency their output was compared with that of the reference accelerometer. The gain on the A/D converter box was adjusted until the two readings were within 0.1% of each other. The second stage was static calibration, which can be easily done. Each accelerometer was held with its bottom side facing a smooth horizontal surface. The gain on the A/D converter box was again adjusted until the reading was reasonably close to the

specified value by the manufacture of 0 volts. It was then turned upside down, adjust the gain until the reading was within 2 ± 0.02 volts.

4.3.3 Measurement Noise

The estimation method and the simulation results presented in Chapter 4 assume that the measurement noise to be random, additive white noise. However, this is usually not true in practice. An analysis of some properties of the measurement noise associated with the transducers used in the test is performed in this section. The properties of the recorded signals and their frequency contents as well as the deviation and mean values. Two cases are considered. The first case is when the hydraulic pump is turned off, and the second when the hydraulic pump is on. With no input signal to the test rig, the signals were captured at a constant sampling rate of 1000 Hz, and filtered using a lowpass Butterworth filter with cut-off frequency 500 Hz. The standard deviation and mean values of the signals from the transducers are presented in Table 4.3.1.

Hydraulic off	S1	S2	S3	S4	S5	S6	S7
mean	-0.242	0.009	-0.174	2.0554	-0.012	-0.017	-0.013
standard deviation	0.012	0.014	0.014	0.0589	4×10^{-6}	37×10^{-6}	9.8×10^{-6}

Hydraulic on	S1	S2	S3	S4	S5	S6	S7
mean	-0.233	0.007	-0.159	2.0527	-0.012	-0.017	-0.013
standard deviation	0.029	0.027	0.014	0.0587	4×10^{-6}	37×10^{-6}	10×10^{-6}

Table 4.3.1 The standard deviation and mean values of measurement noise: where notation of the variables are given below:

- S1 – ground input acceleration [m/s^2]
- S2 – unsprung mass acceleration [m/s^2]
- S3 – sprung mass acceleration [m/s^2]
- S4 – force from the load cell [kN]
- S5 – ground input displacement [m]
- S6 – unsprung mass displacement [m]
- S7 – sprung mass displacement [m]

The maximum error in the accelerometer reading was ± 0.02 volts, when the signals are converted to acceleration (1 Volt/g) this error becomes readings multiplied by the calibration scale of 1 volt/g , the errors increase to $\pm 0.2 \text{ m/s}^2$. This is a major contribution to the deviations of the recorded mean values of S1-S3 zero value. However, these trends can be removed prior to the estimation procedure or the following analysis in this section. When the hydraulic pumps are turned off, the standard deviation of S1-S3 are small, suggesting an acceptable accelerometer set-up for the test. Greater deviations are introduced when the hydraulic pump is turned on. The mean value from S4 can be considered to be the total weight of the rig seen by the load cell. However, it has a high standard deviation of almost 59 N. The displacement transducers have small standard deviations indicating small variations in the measurement readings.

The frequency contents of the signals are analyzed by observing their power spectral density (PSD); plots of the PSD against frequency when the hydraulic pump turned off are presented in Fig. 4.3.4, and when they are on in Fig. 4.3.5. Furthermore, in order to test whether the signal has a normal distribution, a simple analysis is performed by observing the 'normal probability' plot of the signal. Fig. 4.3.3(a) shows the PSD of a normally distributed random sequence; 1000 elements, the mean and with standard deviation values of 0 and 1 respectively. Fig. 4.3.2(b) illustrates the

'normal distribution' test concept. The elements of the sequence are plotted against the cumulative probabilities of the standard normal distribution, these are displayed with the symbol ' + '. Superimposed on the plot is a line joining the first and third 'quarters' of the sequence. This line is extrapolated out to the ends of the sequence to help evaluate the linearity of the signal. If the signal is normally distributed, the plot will appear linear. Other probability density functions will introduce curvature in the plot.

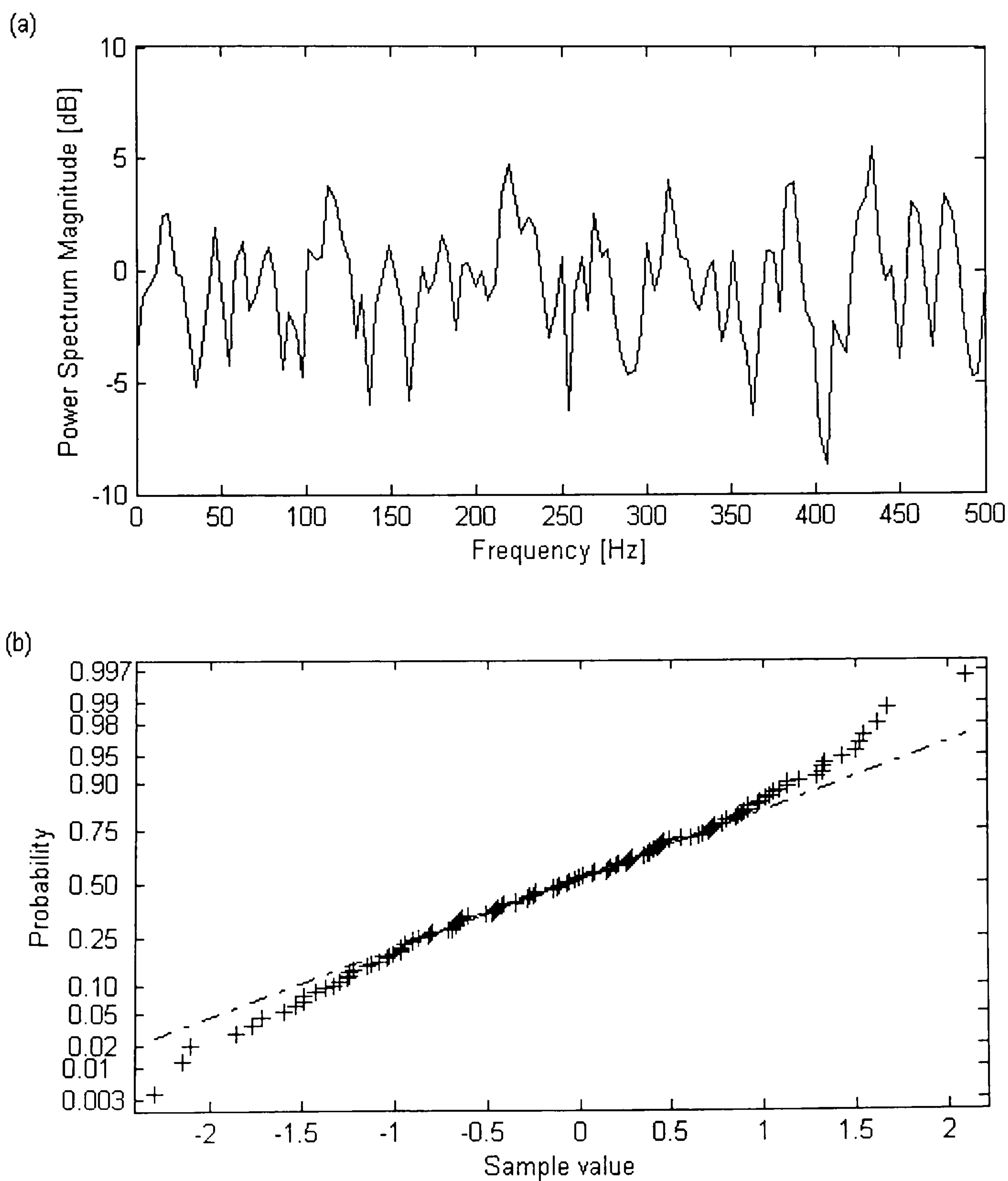


Fig. 4.3.3 PSD of a normally distributed sequence (a), and its probability plot (b).

When the hydraulic pump is turned off, it can be said that the PSD of S1-S3, see Fig 4.3.4(a)-(c), are reasonably flat and similar to that in Fig. 4.3.3(a). From Fig. 4.3.4(d)-(f), a horizontal level may be drawn for each plot. Some irregular peaks with amplitudes of the PSD are higher than the horizontal levels are seen at high frequencies. For the PSD of S4 shown in Fig. 4.3.4(g), there is greater number of high amplitude peaks seen at frequencies over 300 Hz. However, these high frequency components of the signal can be filtered out. As far as the frequency range of interest is concern, the PSD of all signals are acceptably flat and similar to that of the normally distributed random sequence in Fig 4.3.3(a).

When the hydraulic pump turned on, the mean level of the PSD of S1 is noticeably increased due to the fluctuations in the pump, see Fig. 4.3.5(a). The effect is greater at high frequencies. This is also seen, with a smaller effect, in S2. From the PSD of other signals, it is difficult to justify whether fluctuations from the hydraulic system have contributed significantly to the signals.

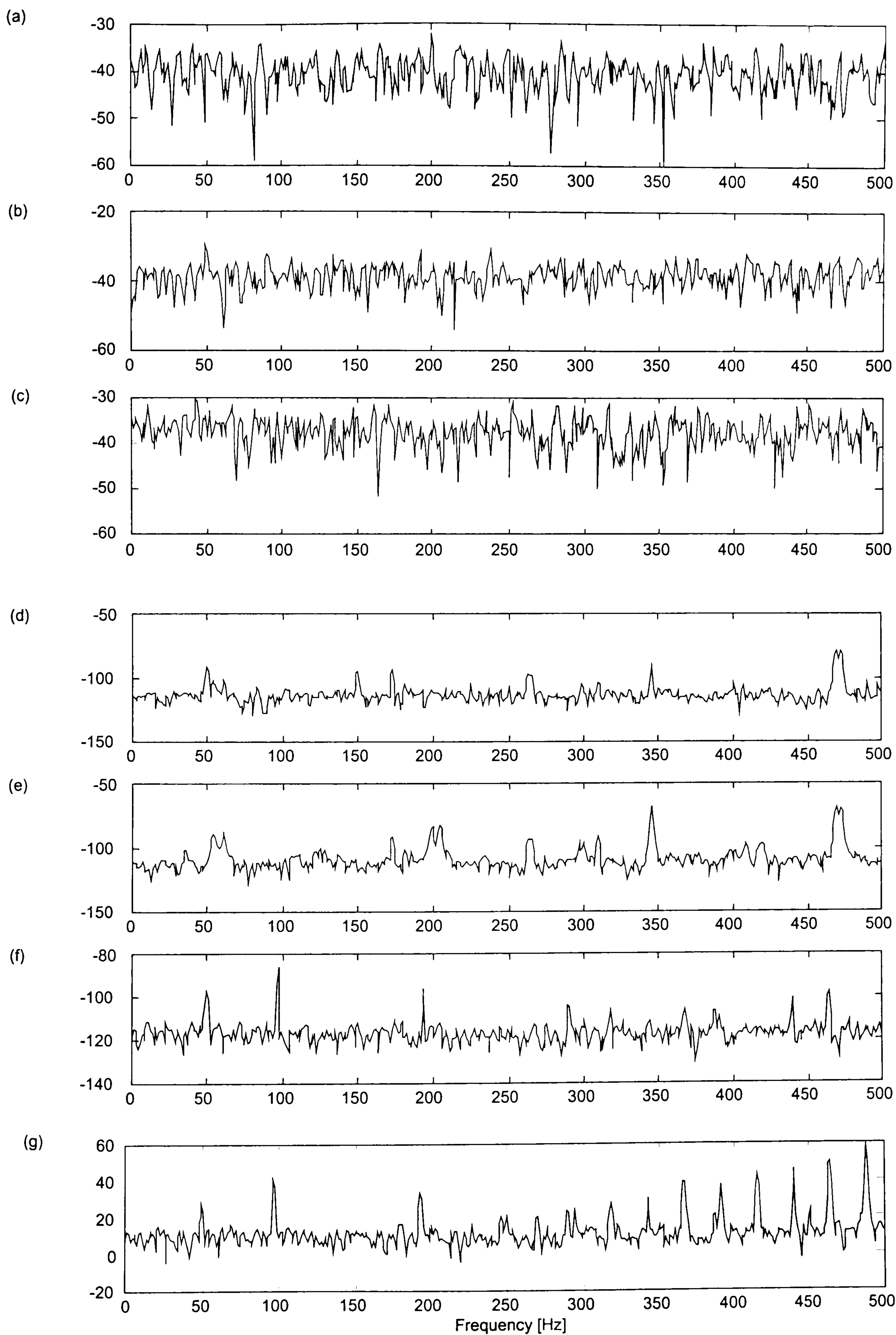


Fig. 4.3.4. Plots of power spectral density [dB] against frequency [Hz] when the hydraulic pump turned off: a) S1, b) S2, c) S3, d) S5, e) S6, f) S7, g) S4.

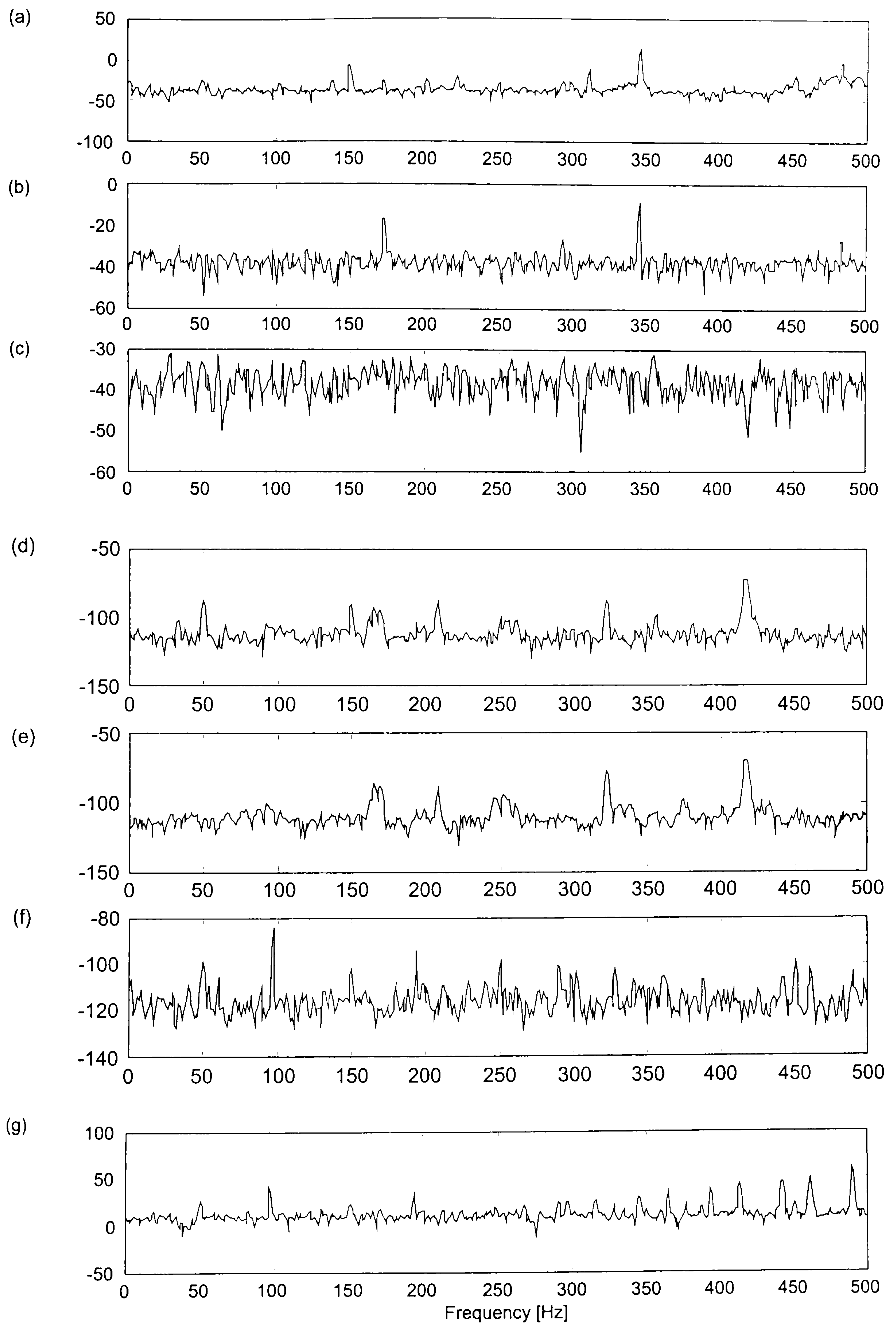


Fig. 4.3.5. Plots of power spectral density (dB) against frequency (Hz) when the hydraulic pump turned on: a) S1, b) S2, c) S3, d) S5, e) S6, f) S7, g) S4.

Another simple analysis was carried out to check whether the signal has a normal distribution, as well as to look for any fluctuation caused by the hydraulic system.

A normal probability plot for each signal is shown in Fig. 4.3.6 when the hydraulic pumps turned off, and in Fig. 4.3.7 with them turned on. Because only the linearity of each plot is of interest, therefore, the element values of each signal, which are on the y-axis, are not included. The range of the y-axis is from the minimum to the maximum values of the signal.

From Fig. 4.3.6(a) –(c), it can be said that the plots are linear. The signals from S1-S3 may therefore be assumed to have a normal distribution. For S5 and S6, see Fig. 4.3.6(d)-(f), small curvature can be observed on the plots. However, it can be arguably that they are linear. For S7, see Fig. 4.3.6(g), the plot is linear within a range, the plot deviates from the straight line at the extreme values.

When the hydraulic turned on, change in shape of S1 can be obviously seen, see Fig. 4.3.7(a). Plots of S1 and S4, shown in Fig. 4.3.7(g) are similar, suggesting there may be a correlation between them. Again, there is no significant change for the other signals. All of the signals were, again, passed through a filter with cut-of frequency of 40 Hz. The resultant plots are shown in Fig. 4.3.8. As seen from the plots, they are all linear, except at their extreme ends. The similarity between S1 and S4, referring to Fig. 4.3.7(a) and (g), disappear as high frequency components of the two signals have been filtered out. This fact is also supports the analysis made on the PSD.

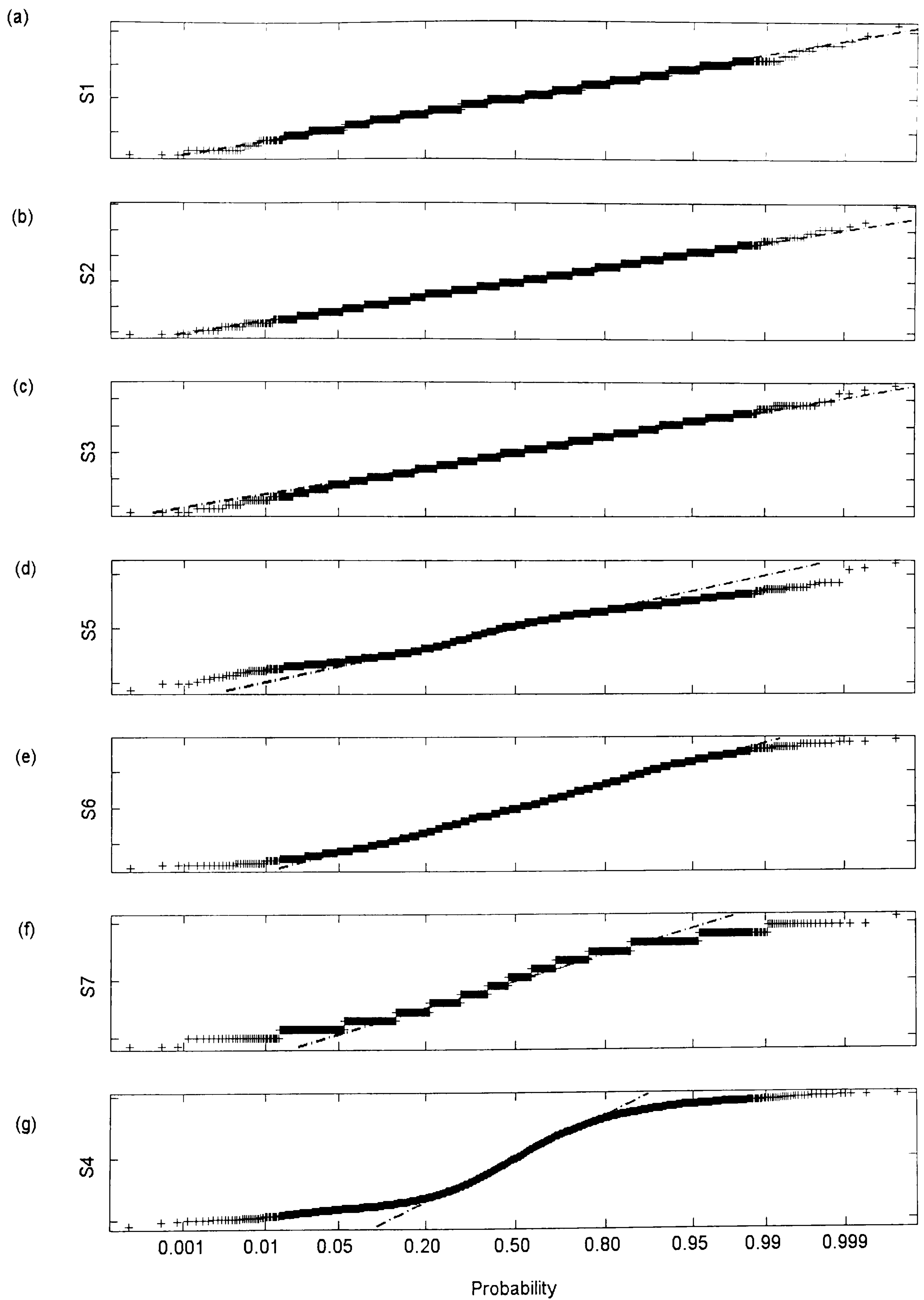


Fig. 4.3.6 Normal Probability plots, hydraulic turned off: a) S1, b) S2, c) S3, d) S5, e) S6, f) S7, g) S4.

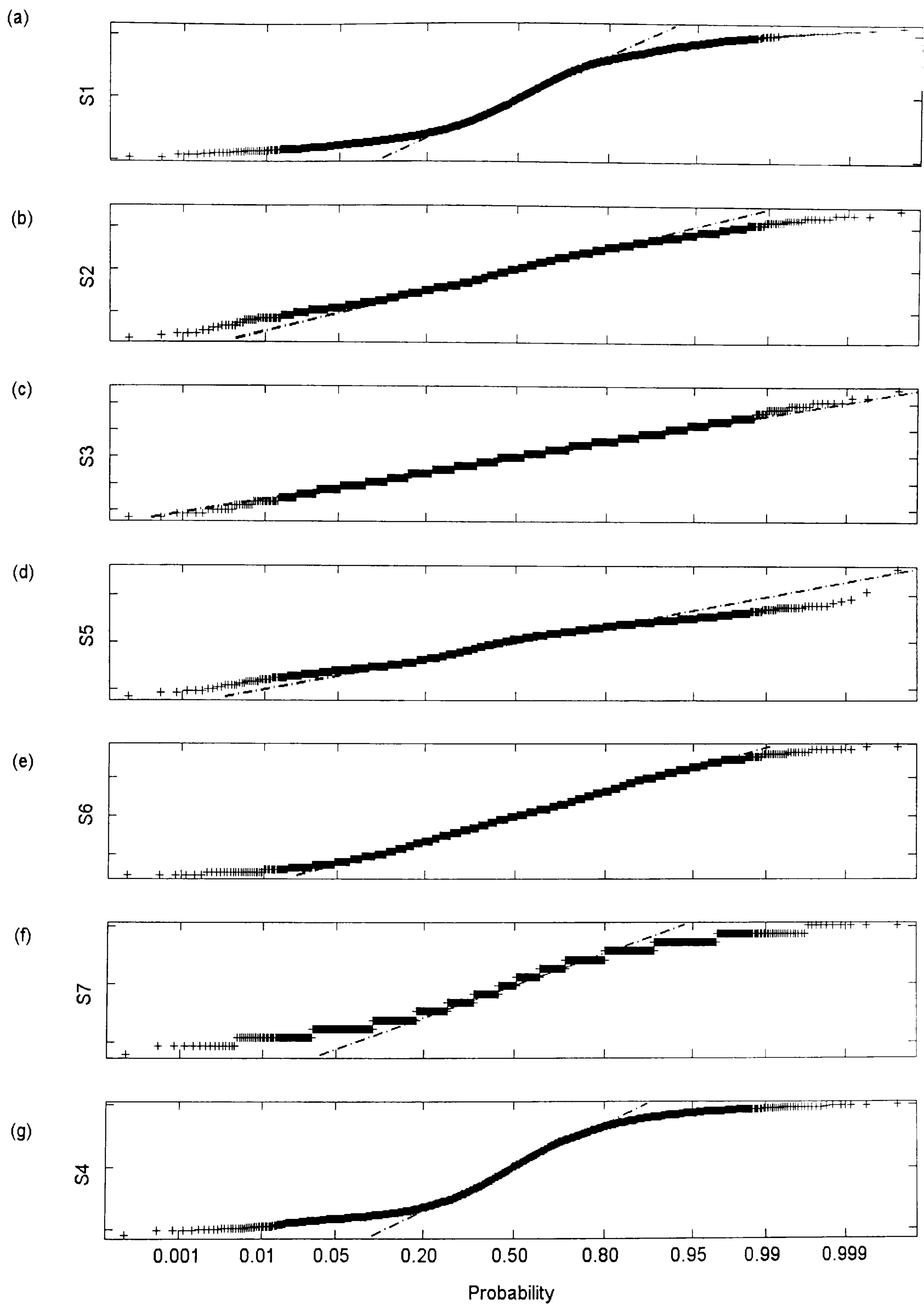


Fig. 4.3.7 Normal Probability plots, hydraulic turned on: a) S1, b) S2, c) S3, d) S5, e) S6, f) S7, g) S4.

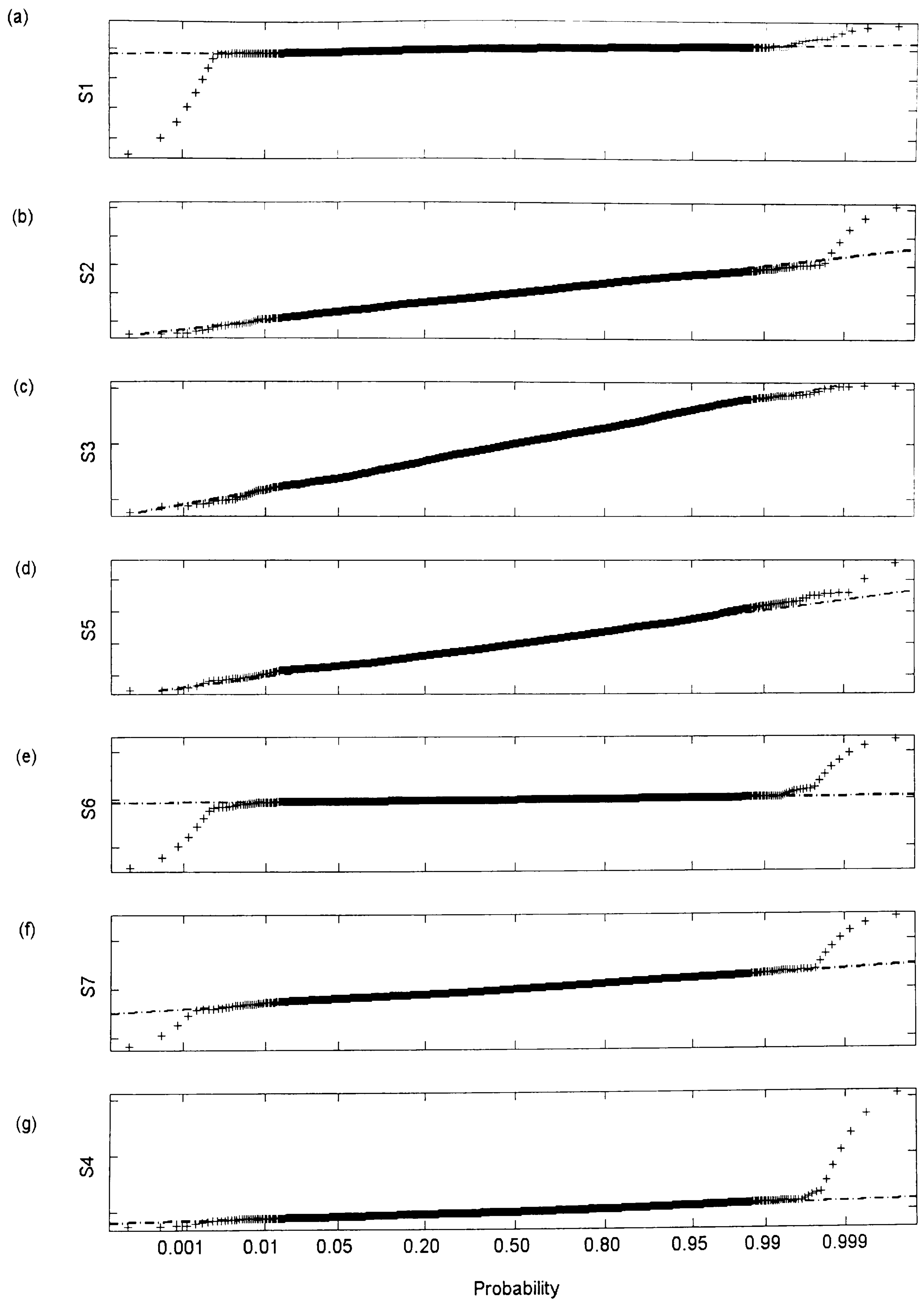


Fig. 4.3.8 Normal Probability plots, hydraulic turned on, filtered at 40 Hz.: a) S1, b) S2, c) S3, d) S5, e) S6, f) S7, g) S4.

The analysis in this section suggests the following important points. Firstly, from the analysis of signal mean and standard deviation, a certain level of confidence can be placed on the transducers used for the experiment. The dispersion level of the signals from the accelerometers and displacement transducers are acceptably small. However, increase in the dispersion is to be expected when the hydraulic pumps are turned on, particularly for the ground input acceleration, S1. Another point to be noted is that the force transducers, S4, has a high standard deviation of about 59 N. This will reduce the confidence on the dynamic tyre force to be used in the estimation method. Secondly, from the analysis of the signal PSD and normal probability plots, the fluctuation produced by the hydraulic system, which is significant at high frequencies, can be filtered out. As far as the frequency range of interest is concern, the fluctuation caused by the hydraulic system contributes only a small effect to the measurement and can therefore be neglected. Finally, the assumption made during the derivation of the estimation method, that is the measurement noise is normally distributed, is justifiably acceptable

4.3.4 Frequency Response

Frequency response test for the single wheel station rig was performed, the input signals for the test were a swept sine wave of frequency 1-30 Hz and decreasing in amplitude. The test responses for the sprung and unsprung mass accelerations for different level of ground input are presented in Fig. 4.3.9 and 4.3.10 respectively.

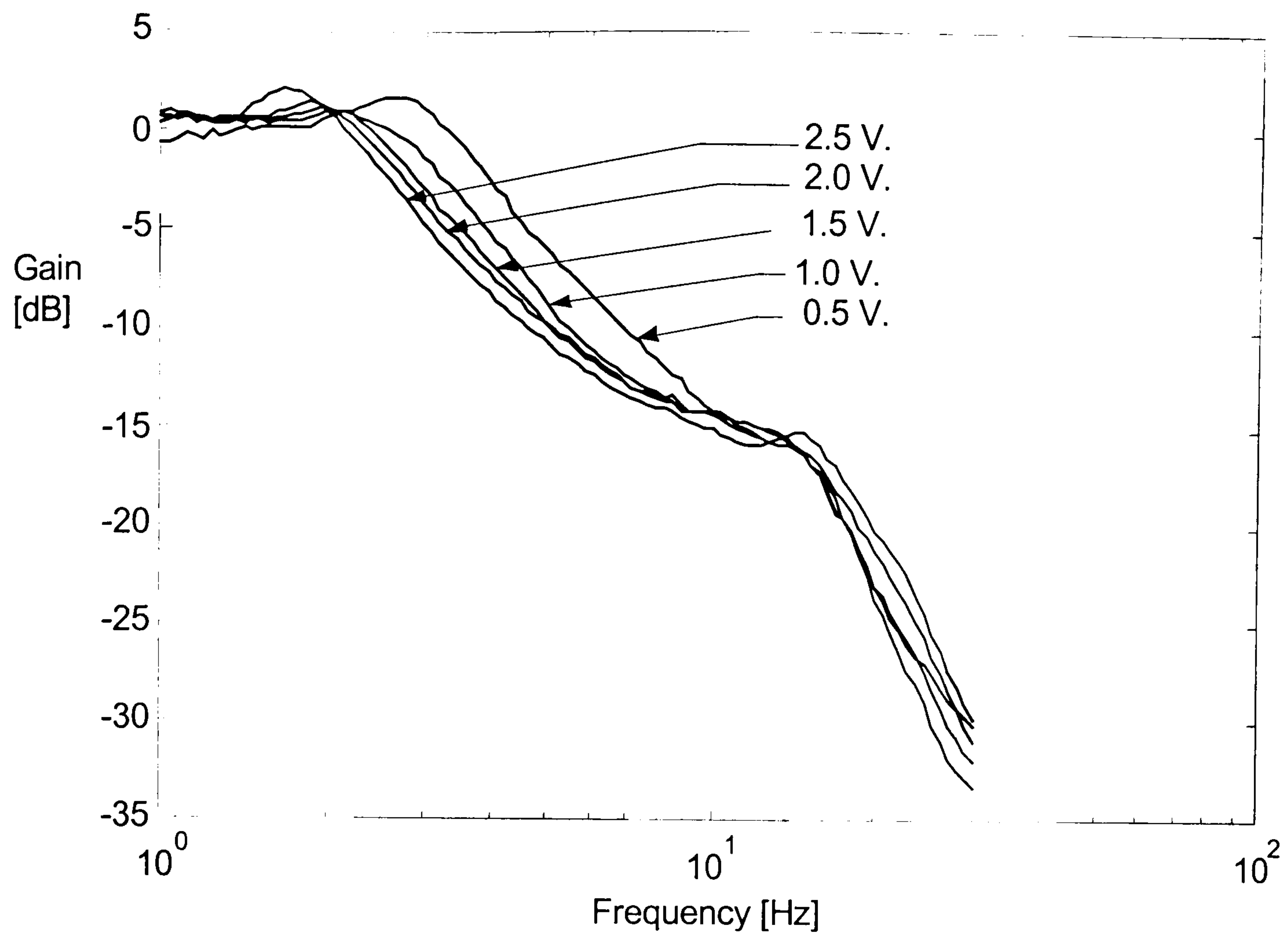


Fig. 4.3.9 Frequency response (sprung mass acceleration/ground input acceleration) for different ground input(RMS voltage).

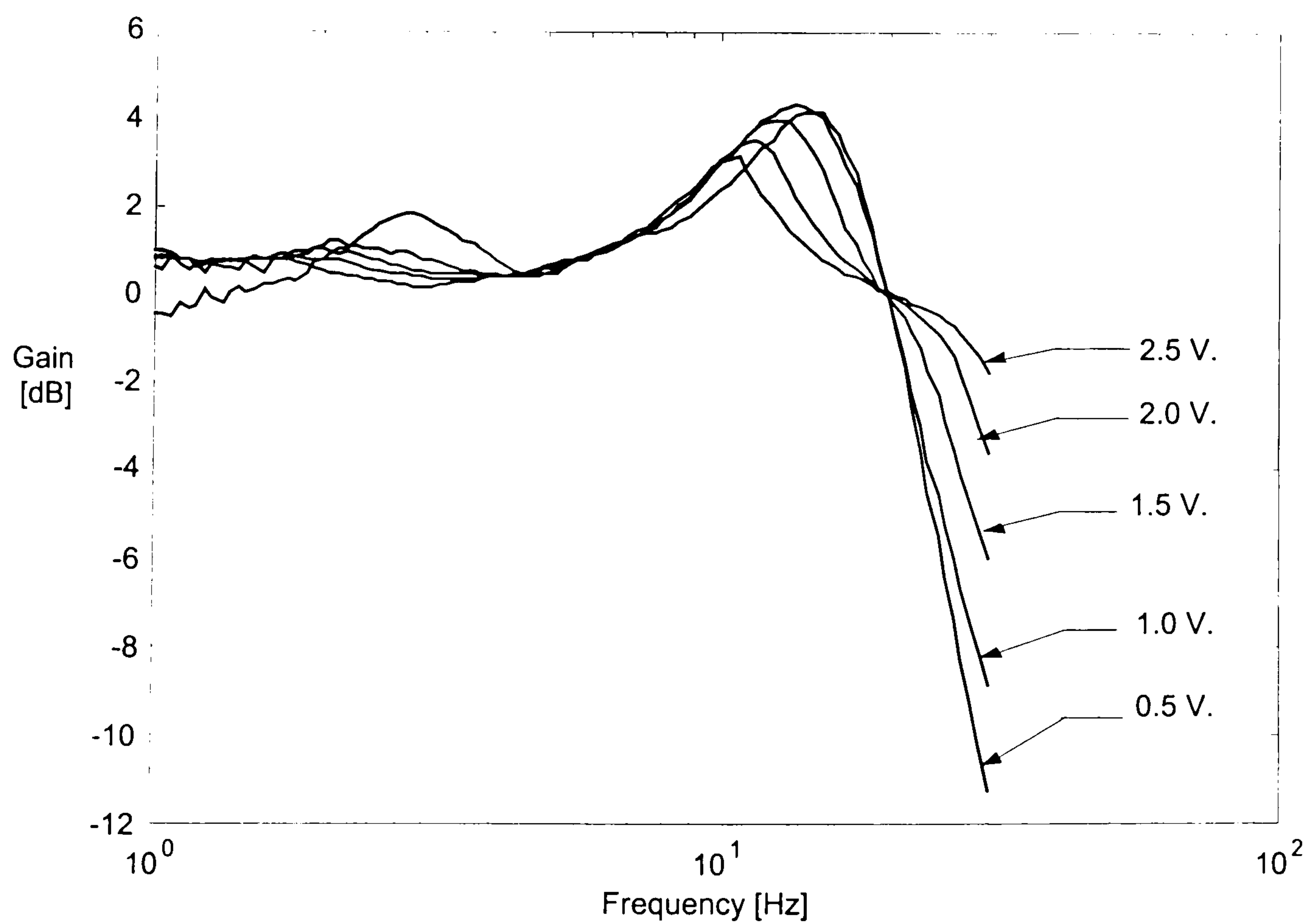


Fig. 4.3.10 Frequency response (unsprung mass acceleration/ground input acceleration) for different ground input(RMS voltage).

As can be seen the system has two modes, the body bounce and wheel-hop modes. The bounce mode is approximately in a frequency range of 1-3 Hz, and 10-15 Hz for the wheel-hop mode. With low amplitude inputs, non-linear friction causes the system resonances to take place at higher frequencies. At low frequencies, it was observed that the suspension components were locked, and the system bounces on the tyre. As the frequency increased, the suspension components broke free and the two masses moved separately. The break point occurred at a lower frequency as the input amplitude increased. This effect can be seen in Fig. 4.3.9 and 4.3.10, where the peaks move in the decreasing frequency direction.

If the data for the frequency responses presented in this section was to be used in an estimation method, for example the least squares or weighted least squares, attempting to estimate the sprung and unsprung masses together with the spring and damping rate of the system, the resultant parameter estimated will certainly depend on which set of data is used. As an example, if the data set used is from a low level input, at the bounce mode we would expect the suspension spring stiffness estimated to be high and sprung mass to be low compared with that of high input. At wheel-hop mode, high tyre stiffness and low unsprung mass are to be expected. However, we would expect the parameters of the system to be the same or reasonably close whether the input magnitude is high or low since it is the same system. To keep the friction effect to a minimum, the input signal was chosen as a swept sine signal with high amplitude at low frequencies and decreasing amplitudes as the frequency increased.

4.4 Parameter Estimation Method

In this section the estimation method employed for identifying the model parameters of the single wheel station is described. The technique used in the time and frequency domains both share a common concept. The technique is depicted in Fig. 4.4.1. First considering the estimation in the time domain when the single wheel station was excited by a swept sine wave input of increasing frequency from 1-30 Hz and decreasing amplitude.

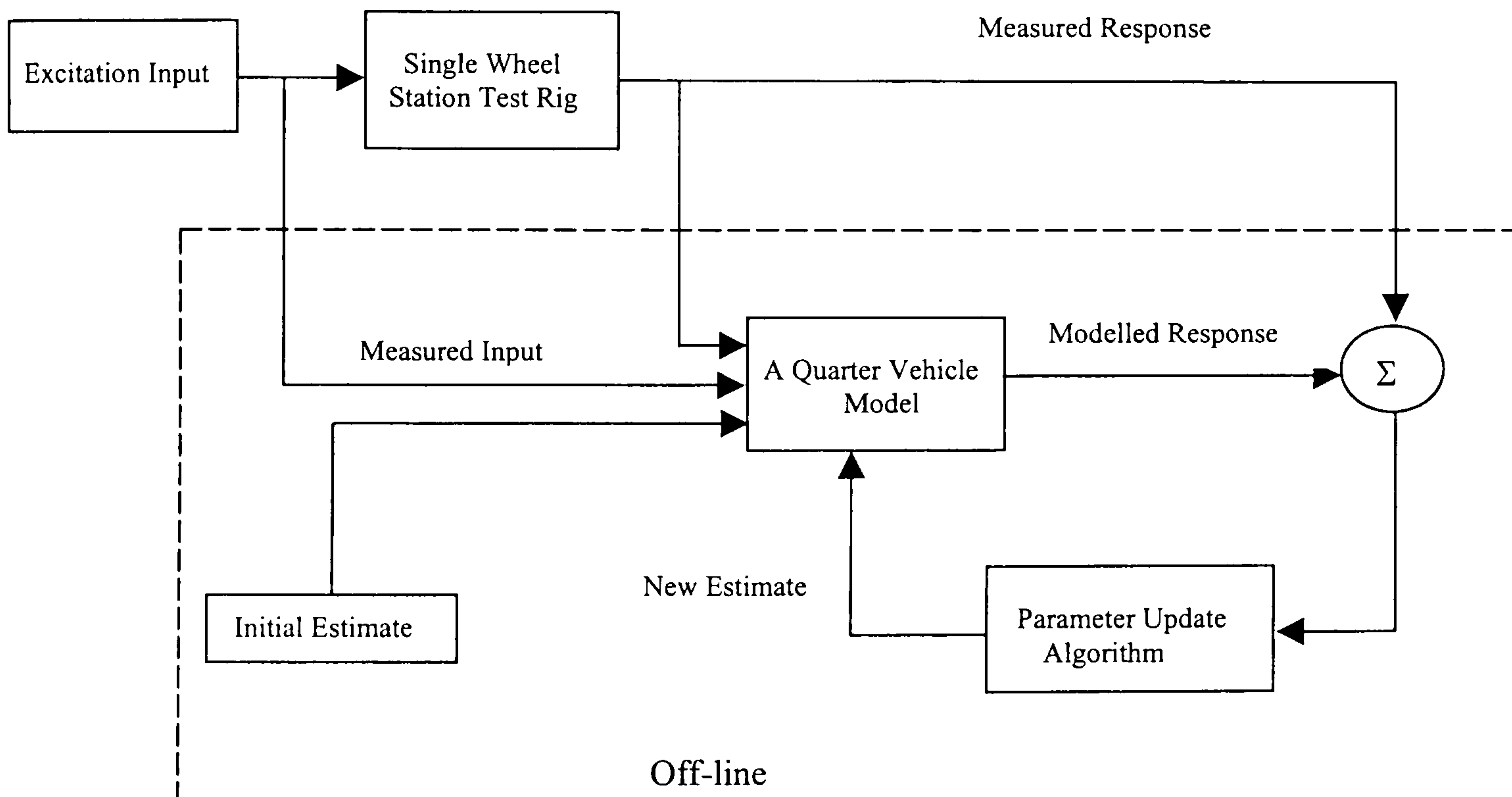


Fig. 4.4.1 Parameter estimation method.

The system responses obtained from the test were the measured time-history signals. They were two signals from the displacement of sprung and unsprung masses, two from the acceleration of sprung and unsprung masses and the tyre force measured by the load cell. It should be noted that the five outputs are simple and economic to measure. These signals together with the measured ground input may be fed to the

model to produce the corresponding model time-history responses. The differences between the measured and modelled responses are then used to formulate an objective function in a parameter update algorithm. The update algorithm in Fig. 4.4.1 refers to a numerical search method that seeks the best possible parameter values by minimising the difference between the measured and modelled responses. Four different numerical search methods were considered; the standard gradient-based (GB) method such as quasi-Newton and Levenberge-Marquardt methods, the Downhill Simplex (DS), the original Differential Evolution (DE), and the proposed dvHDE methods. The dashed box in Fig. 4.4.1 indicates that the search process is an off-line and iterative procedure. Starting with one or a population of initial estimated parameters, the numerical search method updates the estimated parameter and moves toward a better region according to the objective function defined. The procedure is repeated with the new estimated parameters until the algorithm reaches its termination criterion. The final parameters are then taken as the solution of the estimation process.

The normalised mean-square error (MSE) is commonly used as a measure of how well the model outputs agree with the system responses [116]; recalling that a MSE of less than 5.0 percent indicates good agreement while one of less than 1.0 reflects an excellent fit. In this work, we monitor the quality of fit by observing the five MSE values of the five measured and modelled time-history outputs. For each system output, the MSE is defined by

$$\text{MSE}(y_e) = \frac{100}{N \sigma_{y_m}^2} \sum_{i=1}^N (y_{m_i} - y_{e_i})^2 \quad (4.28)$$

where $\sigma_{y_m}^2$ is the variance of the system output y_m . y_e is the corresponding model output and N is number of sampled data. The optimal parameter is then an individual that optimises the five MSEs simultaneously. The problem is therefore a multi-objective optimisation problem, in which a vector of five objectives is to be minimised. In this work, a compromise approach suggested in literatures, see for example [120-122], was employed such that objective function or the fitness for an individual parameter estimate is expressed by

$$J(x, w_k) = \left(\sum_{k=1}^5 w_k^2 |J_k - J_k^*|^2 \right)^{1/2} \quad (4.29)$$

where J_k is the MSE value belonging to the k^{th} model output, J_k^* is the ideal, usually not attainable, zero point, and w_k is weights given to the objectives to emphasis the different degree of importance.

In the estimation in the frequency domain, the measured responses were obtained from the frequency response tests. They were the amplitude ratios and phases belonging to the accelerations of the sprung and unsprung masses relative to the acceleration of the excitation input. The corresponding modelled frequency responses were obtained using the technique developed in section 4.2.3. The parameter estimation was formulated as a minimisation of the weighted sum of difference between the modelled and measured frequency responses, using both the gains (the amplitude ratios) and phases of the sprung and unsprung masses. The weights given to the data at each frequency point were the reciprocal of the variance of the measured data obtained from a small number of repeated experiments. The search procedures

were carried out in a similar fashion as that in the time domain case. All computational codes both the estimation in time and frequency domains were implemented in MATLAB [108]. The estimation results are presented in the next section.

4.5 Estimation Results

This section presents the estimation results obtained by using the method described in section 4.4. It also discusses how well the model performs at mimicing the system being studied. The performance of the four different numerical search methods, the GB, DS, DE and dvHDE methods, are discussed and compared. The estimated parameters of the single wheel station such as masses and spring stiffnesses are also compared against that from the measurement data obtained by measuring each component separately. The results from the estimation in the time domain are discussed in section 4.5.1. The result suggested incorporating the system process and measurement noise information into model. This led to the use of the Kalman Filter, and the resultant model's quality of fit and possible errors are summarised in section 4.5.2. The estimation in frequency domain is then followed in section 4.5.3.

4.5.1 Time Domain Analysis

Having obtained the experimental data required for the estimation method, the first attempt was to estimate the parameters employing a linear two degrees of freedom

(2DOF) model, denoted here as Model 1. The model outputs were obtained according to equation (4.5)-(4.9). The quality of fit in term of MSE values is given in Table 4.5.1. To give an overall visual comparison of how well the model performed, time-history of the measured and modeled outputs are plotted on the same graphs, in Fig. 4.5.1-fig.4.5.5. From the plots, it can be seen that the model is ‘in the right order’, however it only roughly fits the system outputs. The highest MSE value, which belongs to the sprung mass displacement, is almost 11 %. Detailed and careful observations reveal that the modeling errors in several aspects can be used to improve the model. First, as indicated by the MSE value and obviously seen from Fig. 4.5.1, sprung mass displacement is asymmetrical. One side of the response tends to be greater than the other. This indicates that the suspension component behaviour is non-linear. Since the suspension deflection, the relative displacement between the sprung and unsprung masses, is small, the suspension spring is expected to behave in its linear region. Therefor, the asymmetrical effect seen on sprung mass displacement should mainly come from the non-linear behaviour of the damper. Secondly, the model tends to over-produce the force acting on the sprung mass resulting in higher displacement and acceleration, see Fig. 4.5.1 and Fig. 4.5.2. The asymmetry effect can also be observed on the sprung mass acceleration, particularly at frequencies between the two natural frequencies, approximately between time 10–30 seconds on the plot.

	$MSE(z_s)$	$MSE(z_u)$	$MSE(\ddot{z}_s)$	$MSE(\ddot{z}_u)$	$MSE(f_t)$
Model 1	10.90	0.14	8.86	8.37	6.89
Model 2	4.23	0.13	7.59	6.82	5.26
Model 3	3.24	0.12	7.21	6.62	3.80

Table 4.5.1 MSE values from different quarter vehicle models.

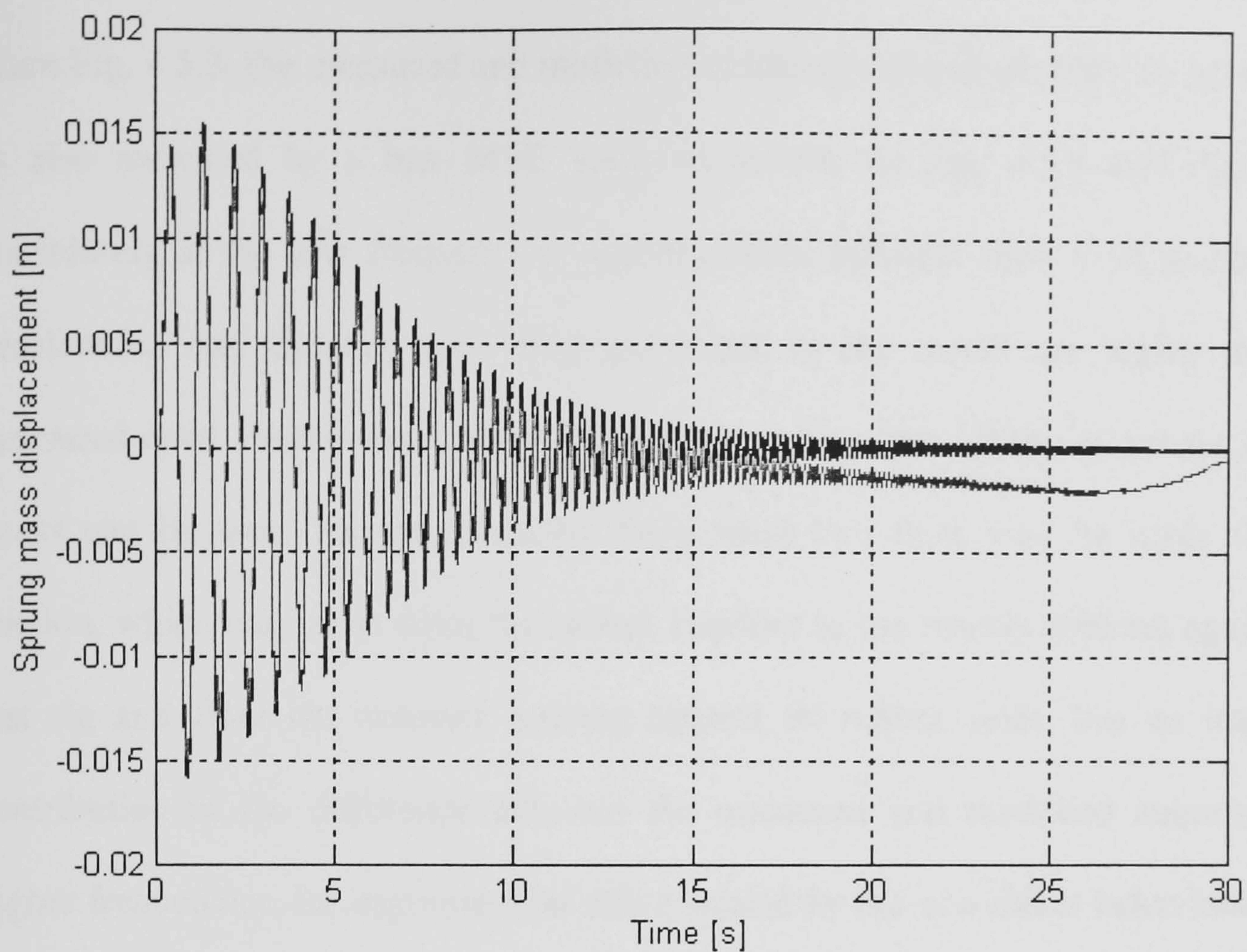


Fig. 4.5.1 Plot of time history of the sprung mass displacement (Model 1):
Modelled - black, Measured – grey.

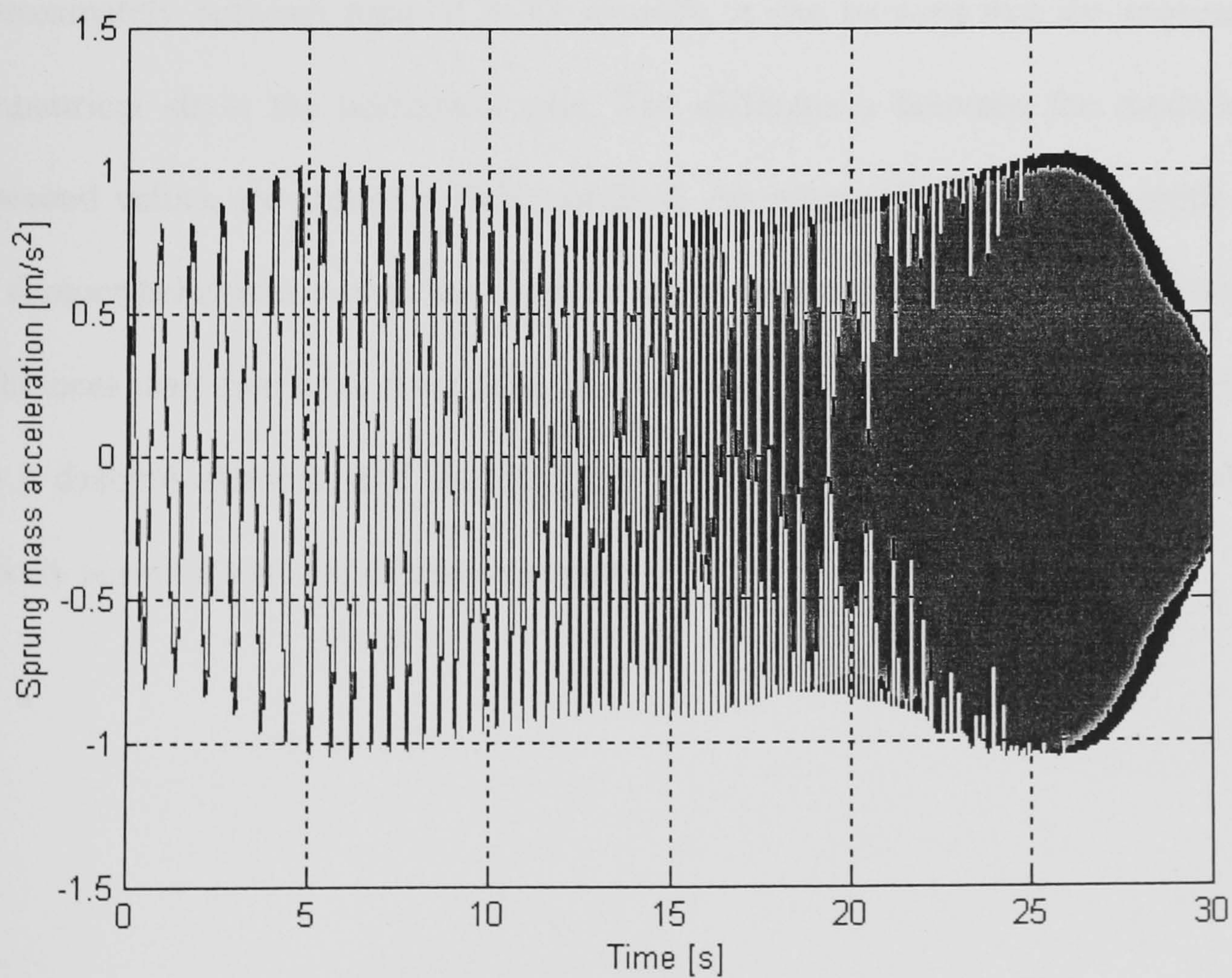


Fig. 4.5.2 Plot of time history of the sprung mass acceleration (Model 1):
Modelled - black, Measured – grey.

From Fig. 4.5.3, the measured and modelled unsprung mass displacements agree well, as also indicated by a low MSE value of 0.14%. In Fig. 4.5.4 and Fig. 4.5.5, considering at the low frequencies, approximately between time 5-10 seconds, the acceleration and dynamic tyre force generated by the model are higher than the measured ones. Furthermore, the non-linearity or distortion of the responses at their peaks can be seen. An interpretation from these two facts may be made that the friction, which may arise from the rollers attached to the masses rubbing against the test rig and from the actuator moving against its rubber seals, has an important contribution to the difference between the measured and modelled responses. At higher frequencies, the asymmetrical effect caused by the non-linear behaviour of the damper has a greater influence on the difference between the measured and modelled responses. Considering both responses at the wheel-hop frequency and thereafter, i.e. approximately between time 31.5–35 seconds, it can be seen that the responses are symmetrical about the horizontal axis. The differences between the modelled and measured values are approximately constant. An interpretation may be made that if the damper behaviour within the range being considered is assumed symmetrical, the differences may then arise from other sources, such as the damper no longer behaving like a dashpot element, and the damping force produced is not only a function of velocity across it but also of displacement.

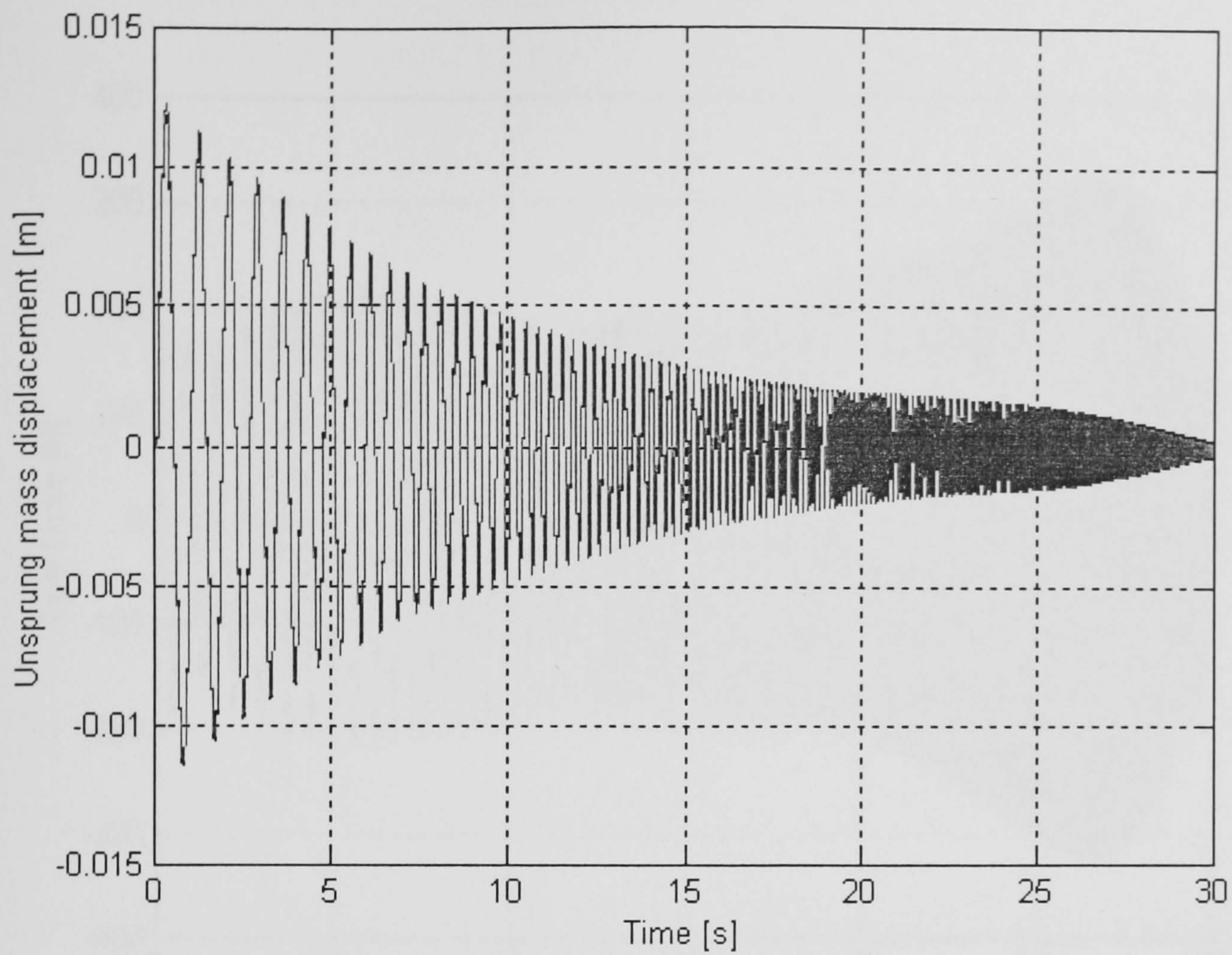


Fig. 4.5.3 Plot of time-history of the unsprung mass displacement (Model 1):
Modelled - black, Measure – grey.

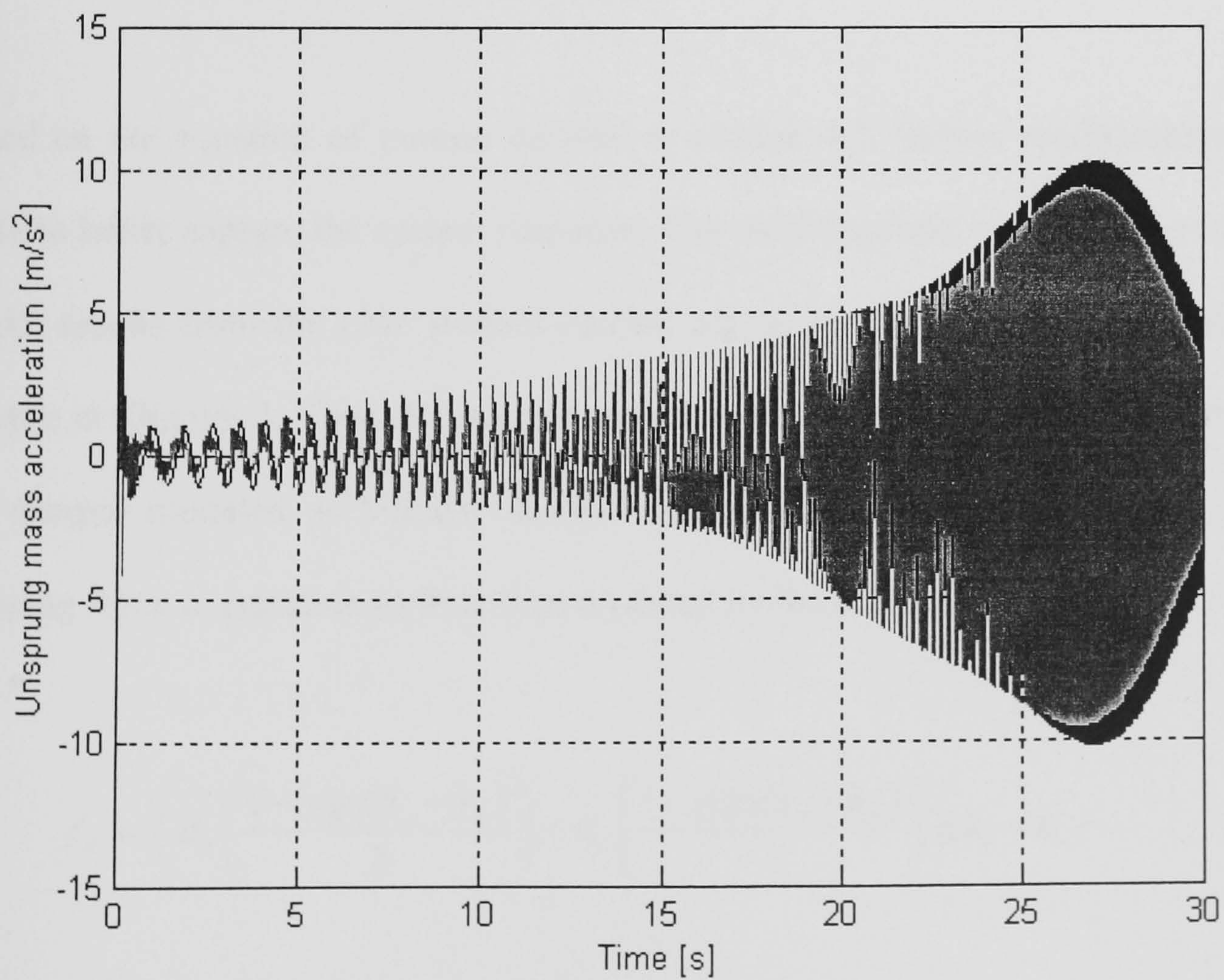


Fig. 4.5.4 Plot of time history of the unsprung mass acceleration (Model 1):
Modelled - black, Measured – grey.

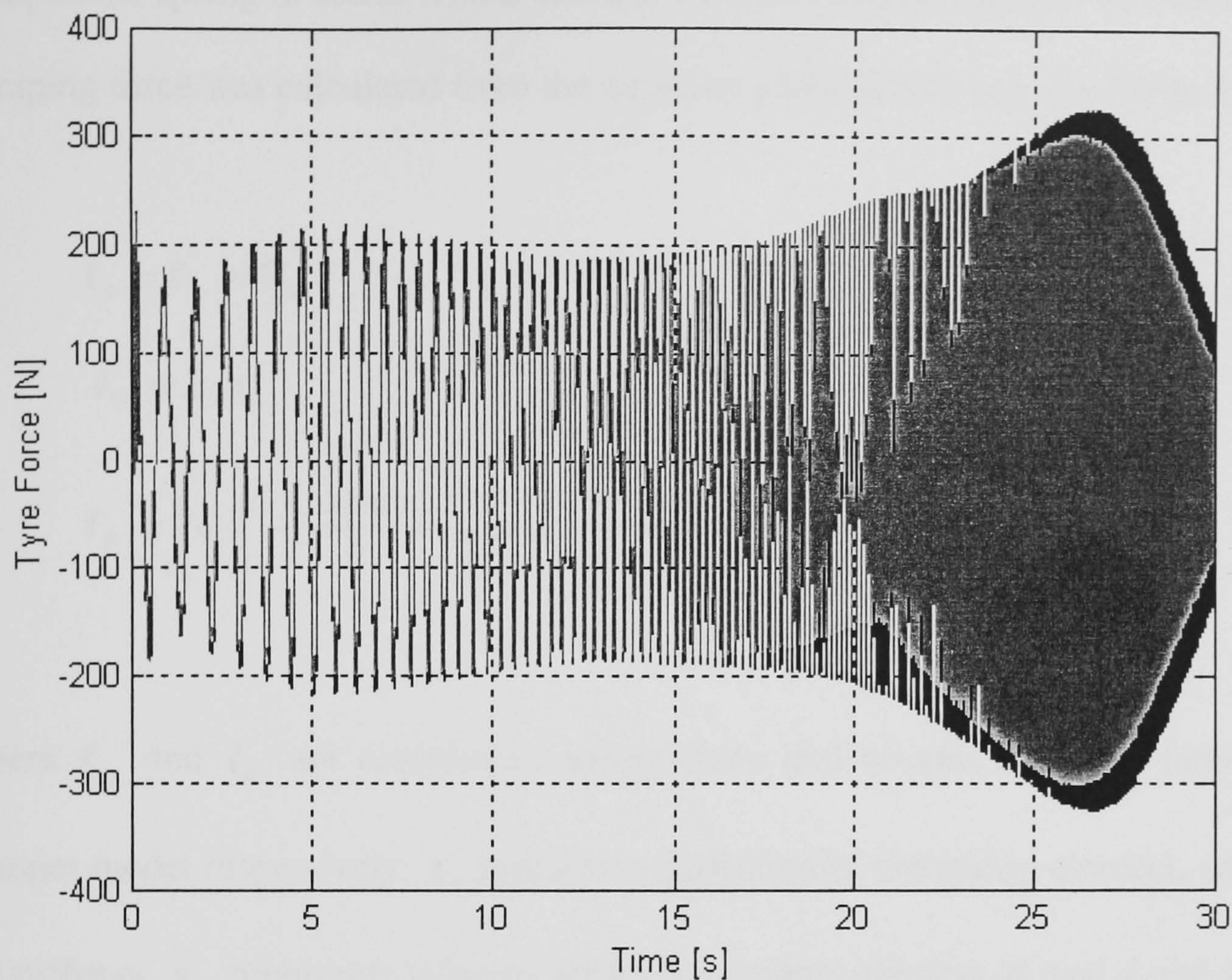


Fig. 4.5.5 Plot of time history of the dynamic tyre force (Model 1):
Modelled - black, Measured – grey.

Based on the equation of motion derived in section 4.2, further modifications were made to better capture the system response. The modifications were mainly suggested by the results from the case studied carried out on the parameter estimation of the damper in Chapter 3. Then, the second model experimented was a 2DOF model with the damper modeled as bilinear dashpot element, denoted here as Model 2. The damping force in equation (4.9) is then replaced by the following equation.

$$f_d = \left(d_r \left(\frac{1 + \text{sign}(\dot{z}_s - \dot{z}_u)}{2} \right) + d_c \left(\frac{1 - \text{sign}(\dot{z}_s - \dot{z}_u)}{2} \right) \right) (\dot{z}_s - \dot{z}_u) \quad (4.30)$$

The third model considered was a 2DOF model with the damper modeled as a linear compliance spring in series with a bilinear dashpot element, denoted as Model 3. The damping force was calculated from the equation given below, see also Table 3.3.1.

$$\mathbf{f}_d = \mathbf{f}_{ds} = \mathbf{f}_{dv} \quad (4.31)$$

$$\mathbf{f}_{ds} = k_{cl} \mathbf{x}_s \quad (4.32)$$

$$\mathbf{f}_{dv} = \left(d_r \left(\frac{1 + \text{sign}(\mathbf{v}_d)}{2} \right) + d_c \left(\frac{1 - \text{sign}(\mathbf{v}_d)}{2} \right) \right) \mathbf{v}_d \quad (4.33)$$

where \mathbf{f}_{ds} and \mathbf{f}_{dv} are compliance spring force and viscous damping force of the damper model respectively. \mathbf{x}_s represents deflection of the spring element, and k_{cl} is its stiffness. \mathbf{v}_d represents velocity across the dashpot element, d_r and d_c are damping rates in the rebound and compression direction respectively. The qualities of fit of the second and third models are given in Table 4.5.1. The model with a bilinear damper gives a better fit when compared with that of the linear model, the MSE values being reduced indicating a better agreement between the measured and modeled outputs. The highest MSE value reduces from about 11% to 7.6%. Further improvement were obtained with Model 3, as indicated by the MSE values. The highest MSE value reduces to about 7%. The highest values of MSE are from the sprung and unsprung mass accelerations. This suggests that they least agree with the measured ones. To provide a visual comparison, time-histories of the responses obtained from Model 3 are plotted in Fig. 4.5.6 – Fig. 4.5.9. Similar arguments made in Model 1 can be said for Model 2 and Model 3. Though Model 2 and 3 were able to capture the measured responses better than Model 1, the discrepancy of the modeled responses from the measured ones can still be seen, particularly at high

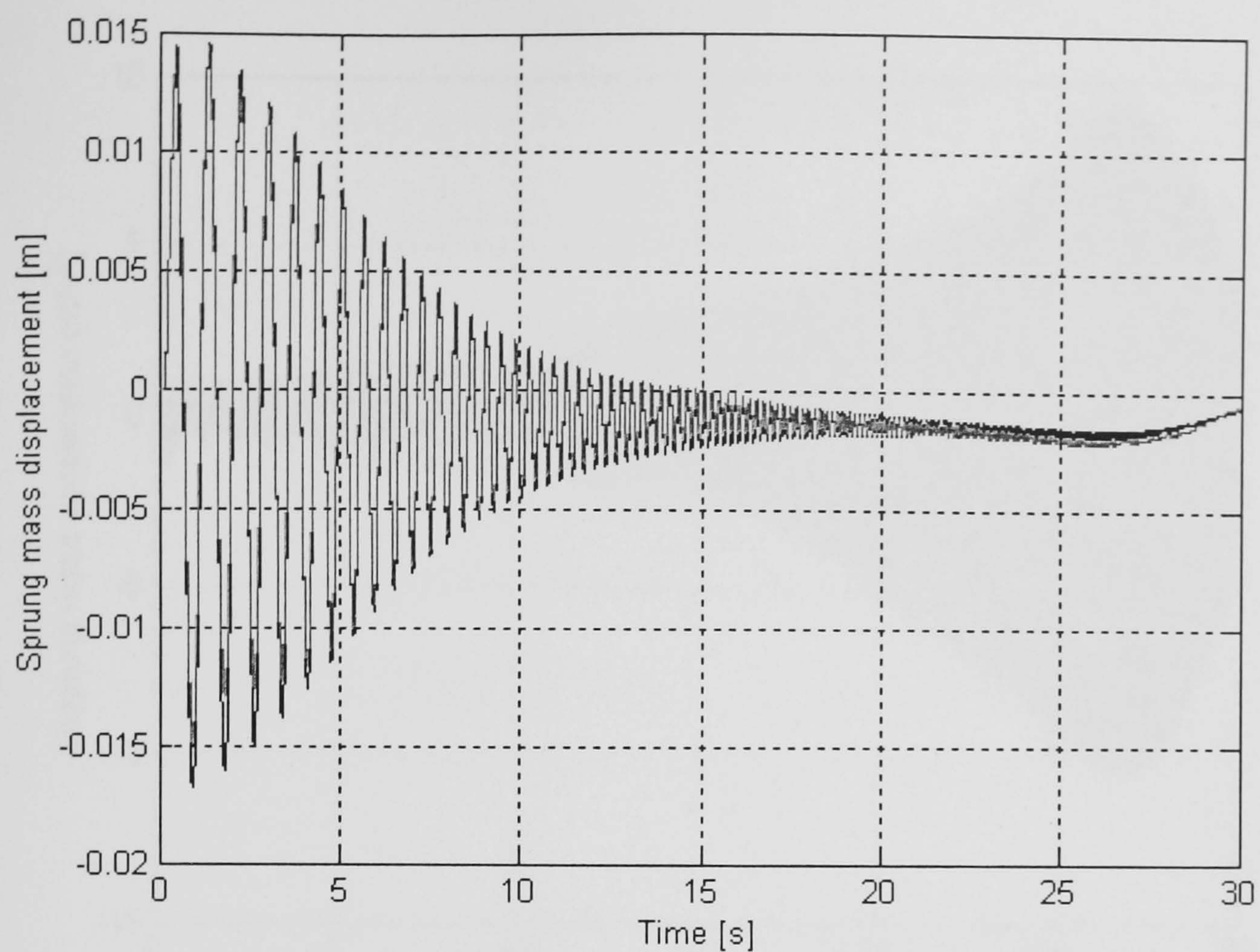


Fig. 4.5.6 Plot of time history of the sprung mass displacement (Model 3):
Modeled – black, Measured – grey.

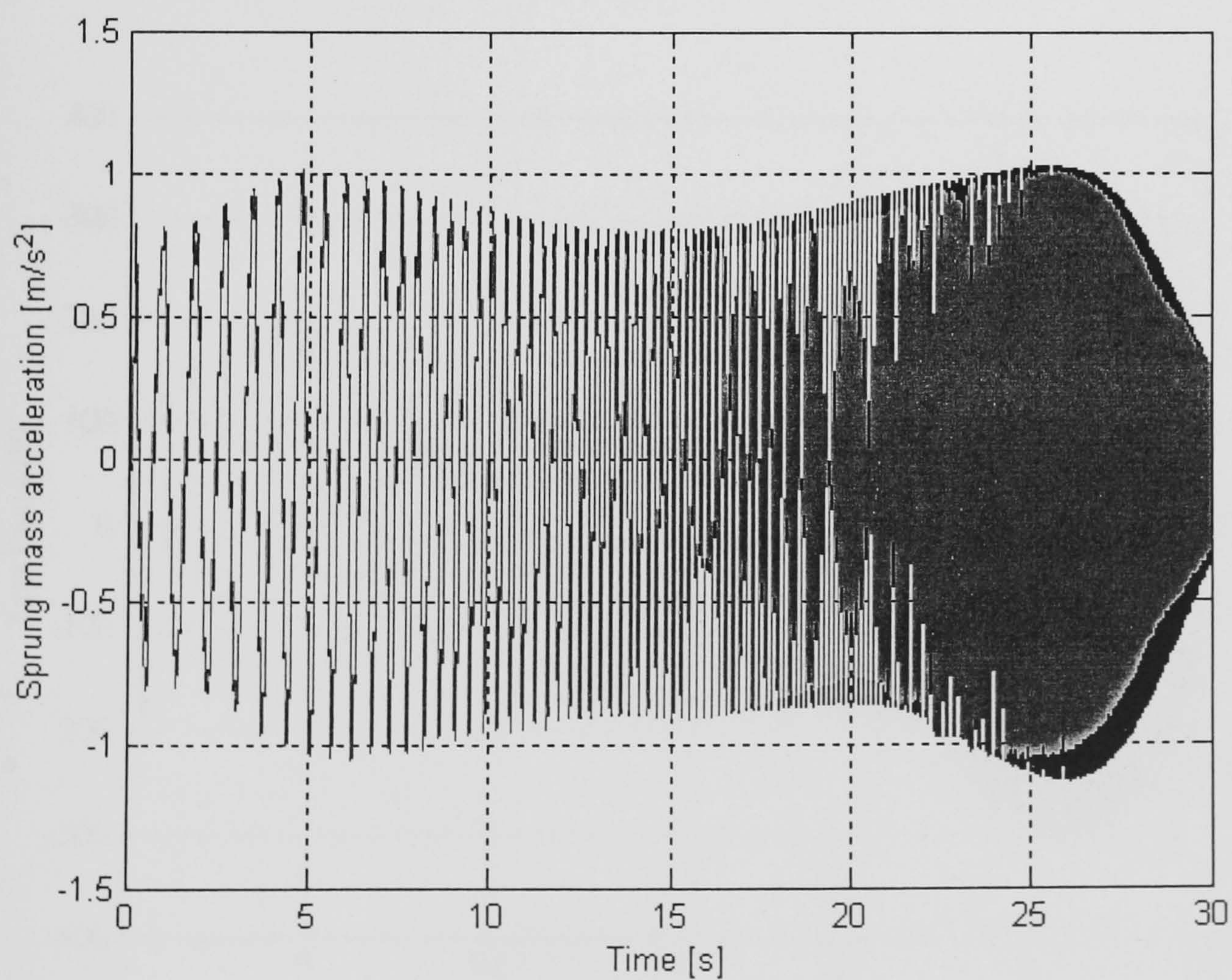


Fig. 4.5.7 Plot of time history of the sprung mass acceleration (Model 3):
Modeled – black, Measured – grey.

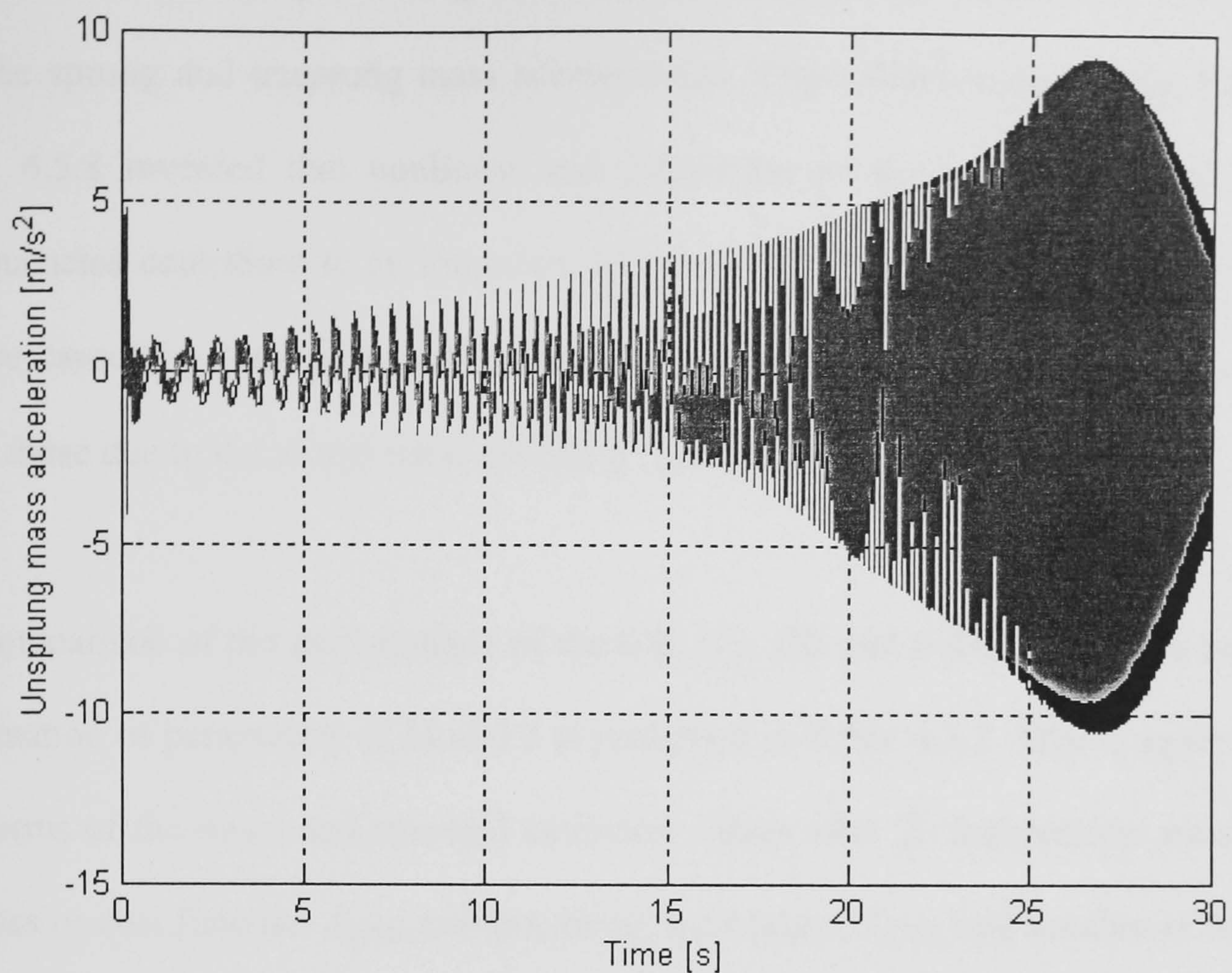


Fig. 4.5.8 Plot of time history of the unsprung mass acceleration (Model 3):
Modeled – black, Measured – grey.

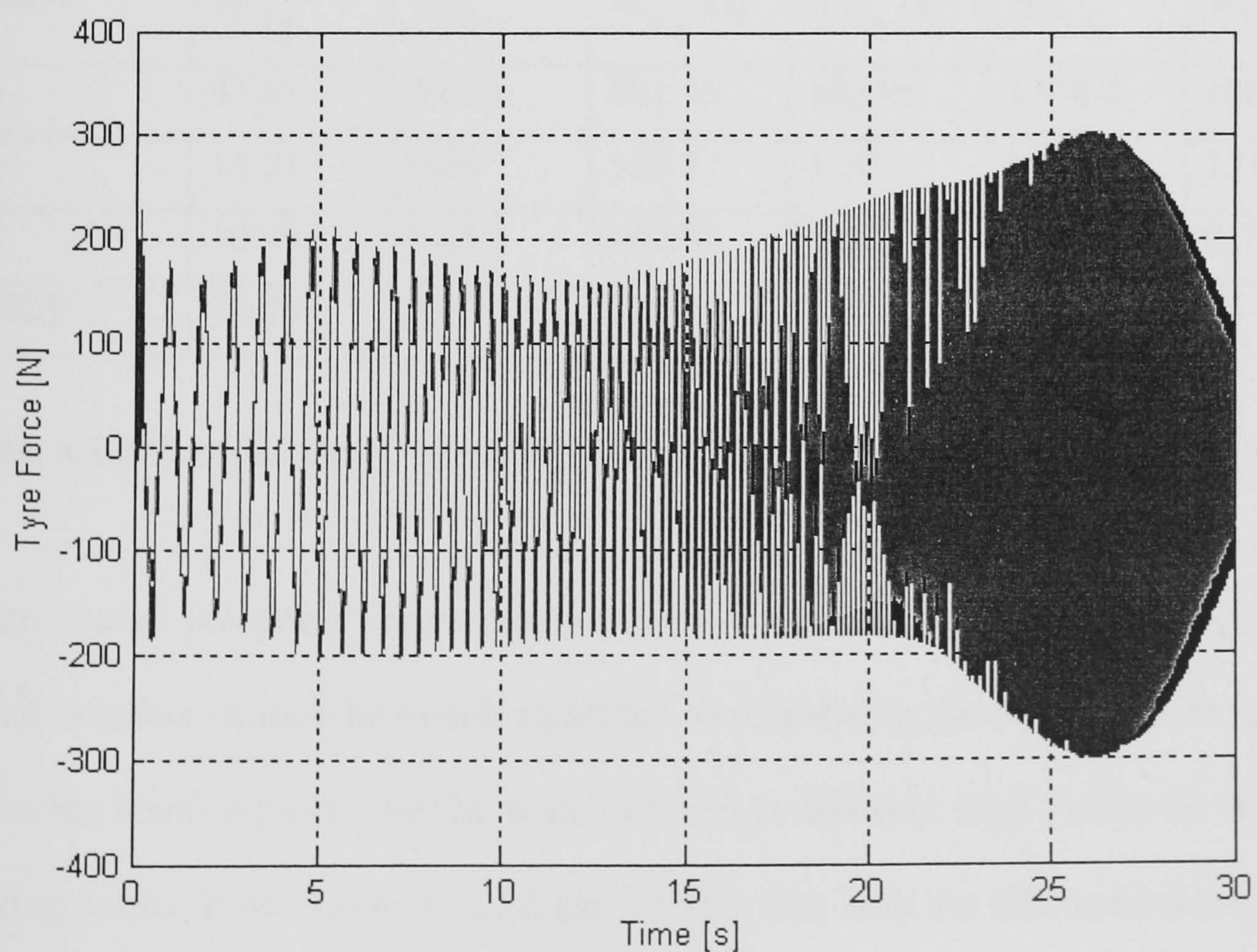


Fig. 4.5.9 Plot of time history of the dynamic tyre force (Model 3):
Modeled – black, Measured – grey.

frequencies, for example see Fig. 4.5.7-Fig. 4.5.9. The worst difference belongs again to the sprung and unsprung mass accelerations. Close observation of Fig. 4.5.7 and Fig. 4.5.8 revealed that nonlinear and distortions of the measured signal at low frequencies contribute to an important additional error to the other sources (such as those caused by ‘non-optimum’ parameters being used to obtain the model outputs, and those due to the model itself not being ‘perfect’).

A comparison of the performance of the GB, DS, DE and dvHDE methods based on estimation of parameters of Model 3 is presented in Table 4.5.2. This is again shown in terms of the mean and standard deviation values after 20 independent runs of the fitness or cost function J_{MSE} , computational time taken, T_{CPU} , and number of function evaluation, N_{fval} .

Method	$m_{J_{MSE}}$	$\sigma_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$\sigma_{T_{CPU}}$ [s]	$m_{N_{fval}}$	$\sigma_{N_{fval}}$
GB	47.61	31.03	421.16	165.95	1174.0	466.9
DS	11.01	0.20	559.52	1.20	1400.4	11.9
DE	10.99	0.13	440.95	6.80	1212.0	6.2
dvHDE	10.88	0.01	334.96	6.52	1031.2	15.1

Table 4.5.2 Comparison of the performance of the GB, DS, DE and dvHDE methods.

There existed infeasible search regions where chromosomes caused the model to be unstable and/or caused the search algorithm to stop during the runs. The GB required a feasible starting point, and the search arrived at different final values for different starting points. From Table 4.5.2, it can be seen that with the GB method the search was trapped in local minima as indicated by high values of the mean and standard

deviation, $m_{J_{MSE}}$ and $\sigma_{J_{MSE}}$ respectively. A high value of CPU-time reflects that when the search enters a valley of the search space, it continues searching in the direction determined by the information of the gradient of the cost function before the termination criteria was met.

The DS method was able to reach the optimum point most of the time, however many times it failed to do so. The final value of the cost function was thus seen to be higher than that of the DE and dvHDE methods. In term of time, the DS method required more time to reach the optimum point than the DE and dvHDE methods. This suggests that the DS method has a slow convergence speed. The results from the DE and dvHDE methods are close to each other. The dvHDE method was able to find the optimum point every run. This was not the case for the DE method, as evident from a slightly higher value of $m_{J_{MSE}}$ and $\sigma_{J_{MSE}}$. The dvHDE performed less number of function evaluations, thus required less time. The results in Table 4.5.2 lead to exclusion of the GB and DS methods, only the DE and dvHDE methods are further compared and discussed in the next sub-section, as they performs better for the problem studied.

4.5.2 Estimation Results with Kalman Filter

In this section the results when incorporating a Kalman Filter into the parameter estimation of the single wheel station are presented. The Kalman Filter provides the ‘unmeasured’ states of the system. These signals together with the measured system outputs are used to obtain the model outputs, which is in turn used in calculating the

MSE values and the objective function of the optimisation process. The MSE values are considered to be a measure of the fitting quality of the model to the five system output signals. The MSE values obtained in section 4.5.1 are relatively high with a peak value of about 7%, bearing in mind that a MSE value of less than 5.0 percent indicates good agreement while one of less than 1.0 reflects an excellent fit. This reflects that the models considered so far are relatively simple, with a small number of model parameters needing to be identified. With these models, modelling and parameter estimation tasks require reasonably short periods of time. The models however have not dealt with friction and other nonlinearities, such as hysteresis which is commonly observed at high frequencies. Other sources that contribute to the difference between the modeled and measured response include issues such as process and measurement noise. The next investigation attempts to further improve the model so that it better mimics the dynamic behaviour of the system being studied. To better represent the behaviour of the system that includes both process noise and measurement noise, the system and output equations may be expressed as follows

$$\dot{\mathbf{z}} = \mathbf{A} \mathbf{z} + \mathbf{B} u + \mathbf{G} w \quad (4.34)$$

$$y = \mathbf{C} \mathbf{z} + \mathbf{D} u + v \quad (4.35)$$

where \mathbf{A} and \mathbf{B} are the system and input matrices respectively, w is process noise, assumed to be Gaussian white noise with known covariance \mathbf{Q} . Matrix \mathbf{G} , which acts as a filter, takes into account the case where process noise is not white. \mathbf{C} and \mathbf{D} are the output and direct transmission matrices respectively and v accounts for measurement noise, again, assumed to be Gaussian white noise with covariance \mathbf{R} . The prior knowledge of the system and measurement dynamic in equations (4.34)-(4.35),

together with the statistics of the process noise and the measurement noise can be used to optimally predict the ‘unmeasured’ system states using a Kalman Filter (KF). For more details about the Kalman Filter, readers are referred to the literature [123-125]. With the KF, available information about the system i.e. the noise properties and the measured system outputs are used in the model building and parameter estimation procedures. The final MSE values when incorporating the KF are given in Table 4.5.3. The values are below 3.5 percent. It can therefore be said that the model give a good fit to the system responses.

The parameters of the single wheel station were measured prior to the investigation, these values are referred to as the component test data, and are listed in Table 4.5.4. It should be noted that the damper compliance stiffness k_{c1} and the two damping rates inserted in the table were obtained from ‘best fit’ for the measured damper force presented in Fig. 4.5.10.

	MSE(z_s)	MSE(z_u)	MSE(\ddot{z}_s)	MSE(\ddot{z}_u)	MSE(f_t)
Without using KF					
Component test data	3.58	0.36	8.51	7.63	3.67
Estimated (dvHDE)	3.24	0.12	7.19	6.62	3.53
Estimated (DE)	3.23	0.13	7.21	6.62	3.80
Using KF					
Component test data	0.001	0.010	4.21	7.70	1.81
Estimated (dvHDE)	0.001	0.007	3.24	2.39	1.68
Estimated (DE)	0.001	0.008	3.28	2.40	1.72

Table 4.5.3 The final MSE values based on component test data and from the dvHDE and DE methods.

The estimated parameters of the single wheel station together with the percentage difference when compared with the component test data are also provided in Table 4.5.4. The dvHDE method produces a very similar set of estimated parameters to that of the DE method, and the differences between the final MSE values of the two methods are very small. Therefore, in terms of MSE values and the estimated parameters, the dvHDE method performs at no lesser quality than the DE method.

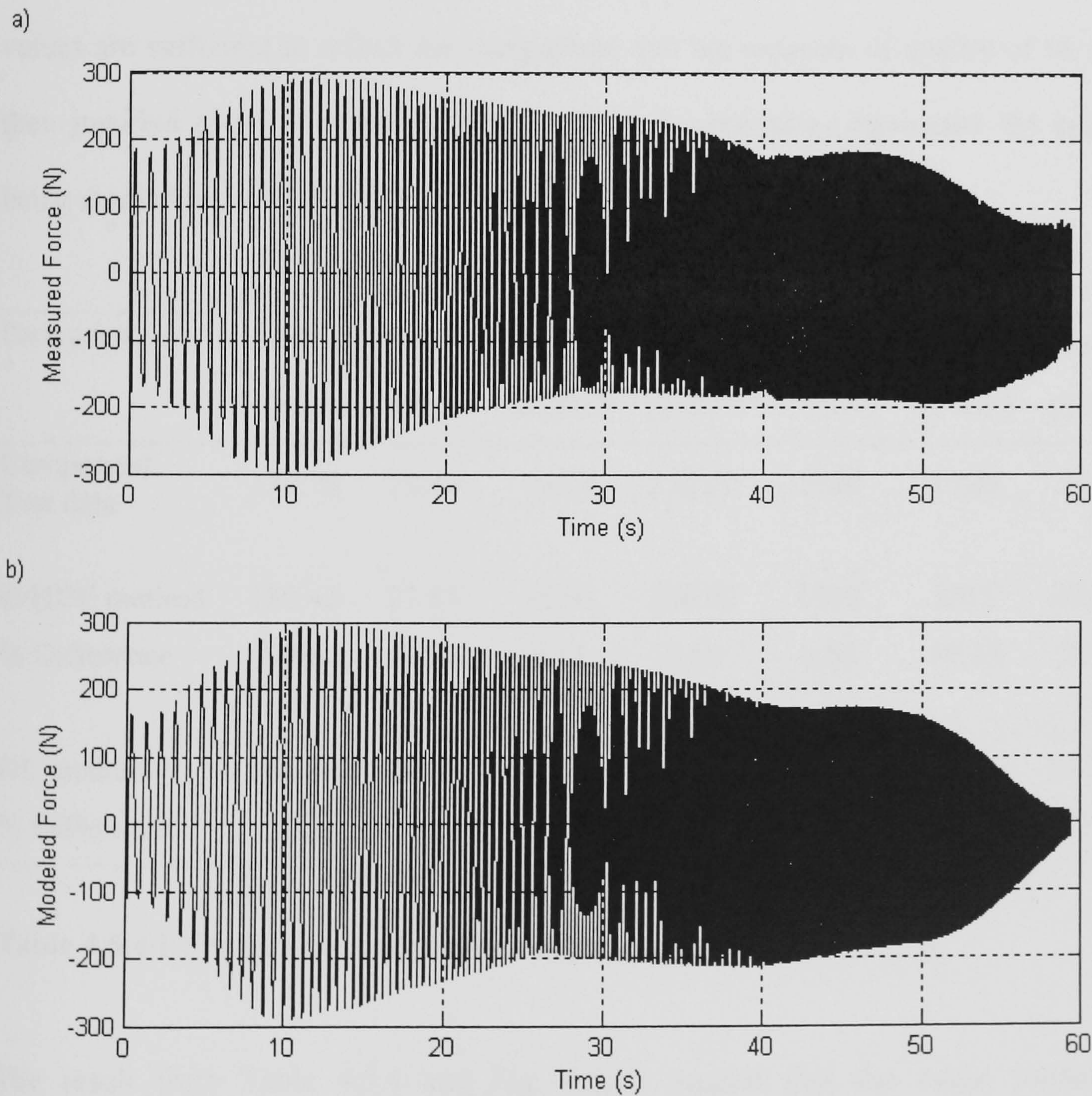


Fig. 4.5.10 Plots of measured and modelled forces of the spring-damper subassembly.

The estimated parameters from the two methods are shown to vary from the component test data, the greatest difference being in the damper parameters. Though, the estimated parameters differ from the component test data, it however, does not imply that the identified model, one with the estimated parameters, is less favourable to the nominal model, one that is based on the component test data. When comparing the MSEs, the values obtained both from the DE and dvHDE methods are all smaller than those that are based on the component test data. A comparison of the time history of the five system responses and that of the identified and nominal models may provide a graphical illustration of this fact. However, they are not show here, as MSE values are sufficient to reflect the comparison and the measure of quality of fit. It is then justified that the model with the estimated parameters represents the system being studied better than the model based on the component test data.

Parameter	m_s	m_u	k_s	k_t	d_r	d_c	k_{d_spr}
	(Kg)	(Kg)	(kN/m)	(kN/m)	(Ns/m)	(Ns/m)	(kN/m)
Component Test data	178.70	29.40	15.01	232.50	1900	1100	348.0
dvHDE method	180.42	27.85	15.93	230.92	1992	1215	438.45
% Difference	0.96	5.27	6.13	0.68	4.84	10.45	25.99
DE method	180.93	27.62	15.84	231.10	2031.4	1271.3	443.51
% Difference	1.25	6.05	5.53	0.60	6.91	15.57	27.45

Table 4.5.4 Estimated and component test parameter values

The result from Table 4.5.4 and Fig. 4.5.10 suggest that the major source of discrepancy between the modelled and the measured responses may come from the spring-damper subassembly. To accurately represent the damper characteristic, a

more complex damper model is needed. It must also be noted that the measured damper characteristic obtained from testing the damper separately on a damper test rig differs from that which may be obtained on the single wheel station in two important aspects. First, in the damper test, where one end of the damper unit is fixed to the test rig, and the other is connected to the input actuator, the information or data obtained does not contain the inertia effect of the damper unit itself. While on the single wheel test rig, the damper unit does move up and down together with the sprung and unsprung masses. This possibly has an effect on the response of the whole system. Secondly, the model employed has not taken account of other nonlinearities in the damper behaviour. The nonlinearities may come from hysteresis, friction between the damper seals and rod. The damper behaviour is also dependent on its displacement, temperature, and input frequency. A closer analysis revealed that at low input frequencies, friction was the main contribution to the difference between the modelled and the measured responses. While at high frequencies, particularly at the high velocity typical of the wheel hop mode, the discrepancy between the modelled and the measured responses is mainly caused by hysteresis. In addition, the non-linearity in the spring-damper assembly is relatively strong since the inclined spring generates a considerable side-loading on the damper, which tends to cause an increase in the hysteresis due to friction between the rod and seal. However, as usually true, the model obtained is a compromise between the computational complexity and a satisfactory description of the system. For the objective of the study in this work, the model proposed has been shown to be sufficient for the purpose. Further improvement in quality of fit, i.e. lower MSE values, may be obtained with a more complex model. This means the model may contain more parameters to be identified, or including other 'unmodeled' effects such as friction from the side-rollers, non-Gaussian noise and

component nonlinearities e.g. cubic spring, and those other effect mentioned for the spring-damper subassembly.

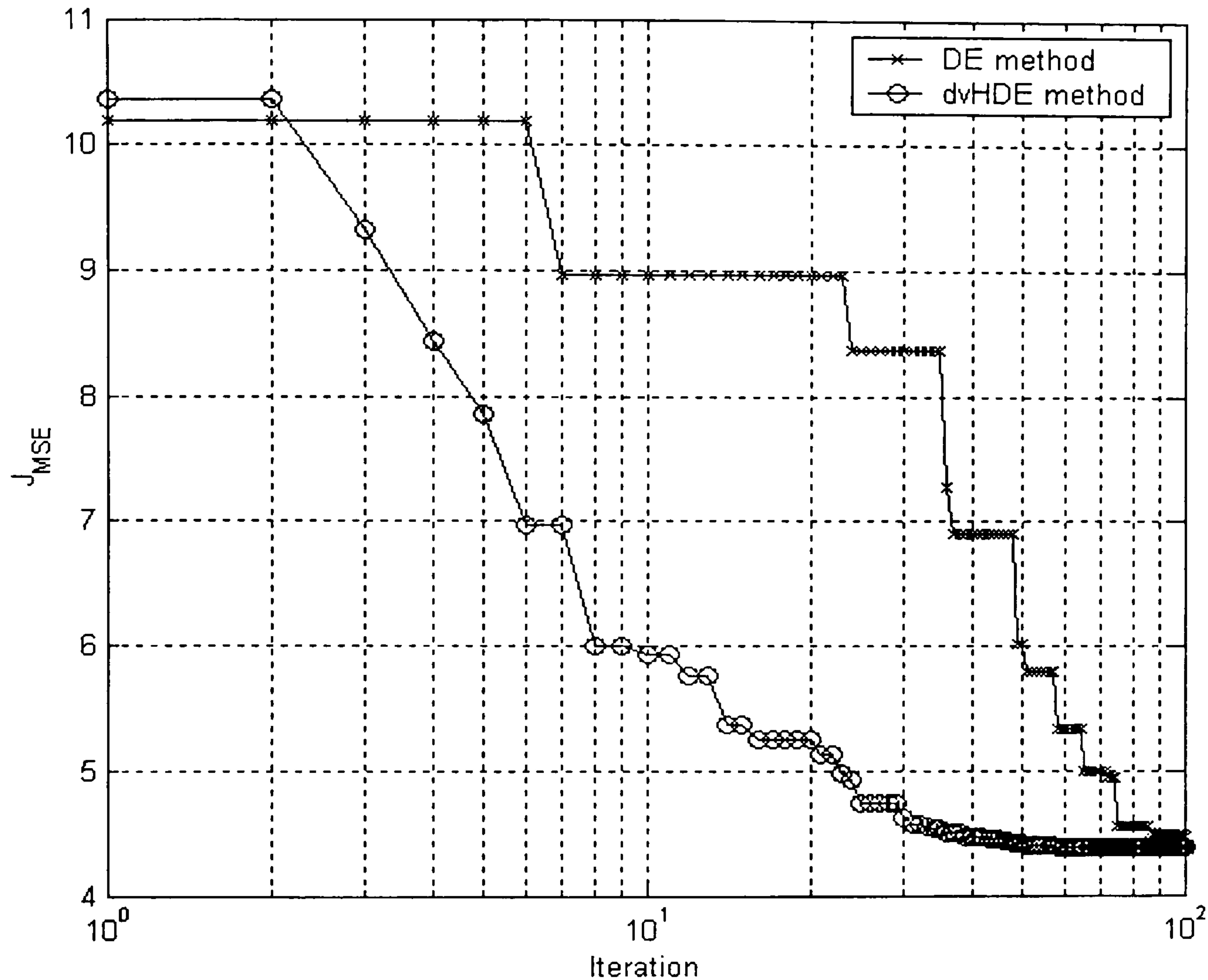


Fig. 4.5.11 Typical plot of fitness values against iteration number, obtained from Model 3 using the DE and dvHDE methods.

Comparing the performance of the numerical search methods, typical plots of the fitness values belonging to the best individuals from each generation, and the corresponding population diversity are illustrated Fig. 4.5.11 and Fig. 4.5.12 respectively. Between the two methods, it is observed from Fig. 4.5.11 that the DE shows a slower convergence rate than that of the dvHDE method. Furthermore, there are several consecutive iterations that the DE method failed to update the fitness value J_{MSE} . As an example, considering an interval from the sixth to twentieth generation, where the fitness values by the DE method in Fig. 4.5.11 remains unchanged and the

corresponding population diversity in Fig. 4.5.12 is decreasing. This situation suggests that the mutation and crossover operations in the DE method only improved the search through selection, where better chromosomes replaced the less fit ones. It however failed to update the best chromosome. As the population further evolved, the fitness value J_{MSE} and the population diversity further reduced. When arrived at one point during the search, approximately at the eightieth generation, the mutation and crossover operations was unable to produce any better offspring chromosomes to replace their parent chromosomes. This is evident on Fig. 4.2.11 and Fig. 4.2.12 as is that both the J_{MSE} and population diversity remained almost unchanged. The DE method then terminated as the maximum iteration was reached. It, therefore, failed several times to reach the optimum point.

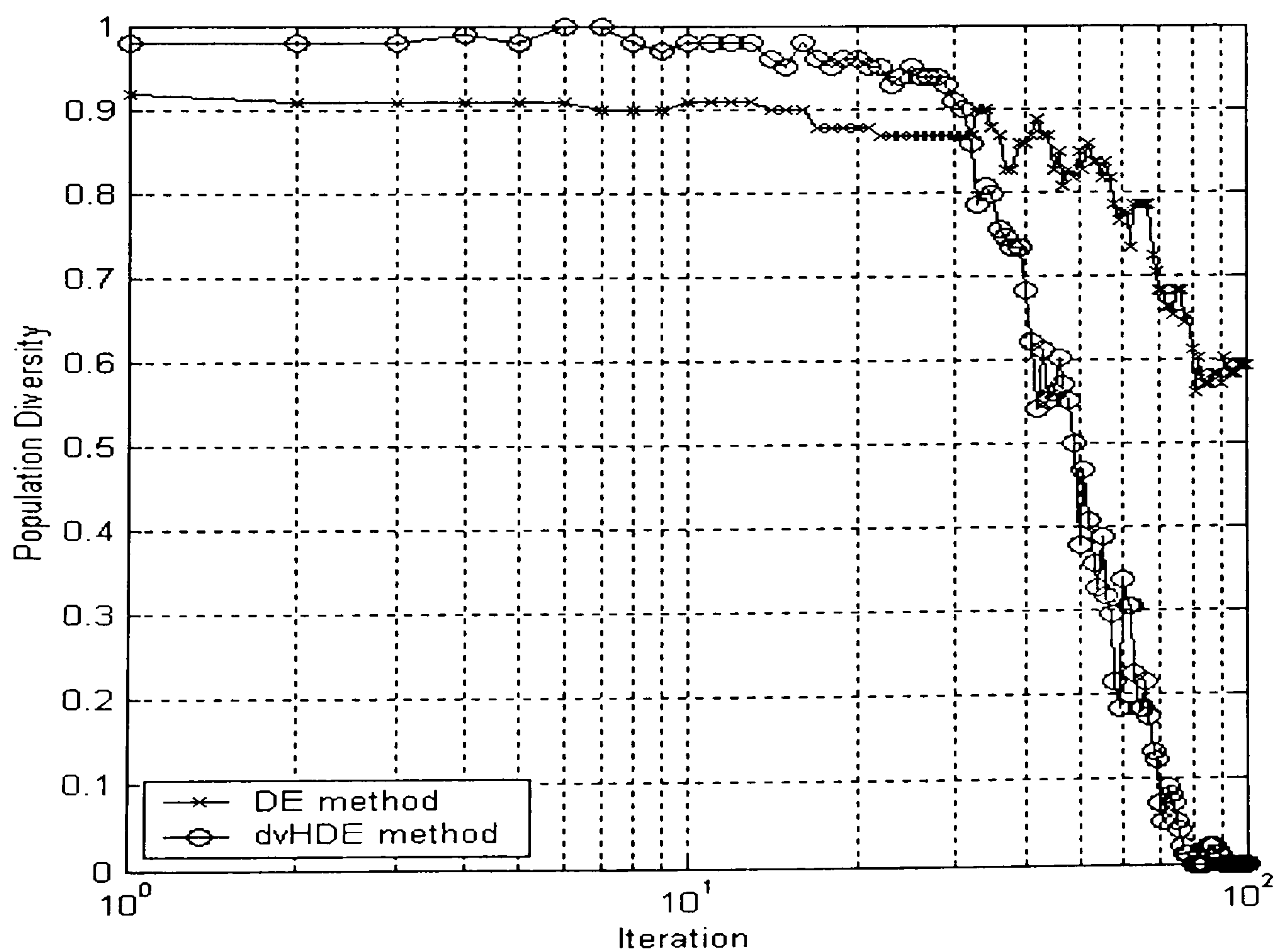


Fig. 4.5.12 Typical plot of population diversity against iteration number, obtained from Model 3 using the DE and dvHDE methods.

The dvHDE method quickly moves the search to appropriate regions, the fitness values successively reducing. After approximately the thirtieth generation and up to the ninetieth generation, the population diversity rapidly decreased and the fitness values approached the optimum value. After that, the population diversity was very close to zero and further decrease in the fitness values were only small, as the iteration reaching the termination criterion set. This suggests the dvHDE method has located and found the optimum point. The average value of the fitness function, $m_{J_{MSE}}$, from 20 independent runs is 4.43 with standard deviation $\sigma_{J_{MSE}}$ of 0.15. While for the DE method, the two values are 6.82 and 3.56 respectively. It can therefore be said that the dvHDE method was able to find the optimum point, and was more consistently able to do so than the DE method. The mean CPU-time, $\alpha_{T_{CPU}}$, from 20 independent runs is 432.45 seconds. Comparing with $\alpha_{T_{CPU}}$ of 533.87 seconds by the DE method, when the search converged, the dvHDE thus improves the convergence by approximately 19 percent.

4.5.3 Estimation Results: Frequency Domain

This sub-section presents the estimation results when the parameters of the single wheel station were estimated using frequency response data as described in section 4.4. These are based on Model 3 from the previous sub-section, and the model was further modified to take into account the friction and hysteresis. The friction was accounted for by a simple element that produced a friction force as a function of velocity, while the hysteresis was accounted for by an element that generated force as

a function of frequency and velocity across the suspension. The estimation results when the excitation input was the swept sine wave of frequencies 1-30 Hz and maximum amplitudes of 7.5 and 12.5 mm respectively presented in Fig. 4.5.13 and Fig. 4.5.14.

From the plots, it can be said that overall the model captured the system dynamic relatively well, though the difference between the modelled and measured frequency responses can still be seen, particularly toward the high frequencies. There were three possible main sources of error that contributed to the differences. First was the limitation of the model, this means the modelling error including the additional elements representing friction and hysteresis. The second was error due to the limitation of the method of obtaining the frequency response itself. The third was due to the errors in the parameter estimation procedure when formulating the objective function. The weights given to the data at each frequency point may cause a possible fault in the cost function. For example, the unreliable points such as at the beginning and the end of frequency range could cause the objective function to be weighted in their favourite. When the parameter search evolved the estimated parameters were formulated according to the objective function, however this was not a serious problem, as seen from the result shown on the plots. For the investigation carried out in this chapter, the model and the parameter estimation procedure have satisfactorily served the intended purpose.

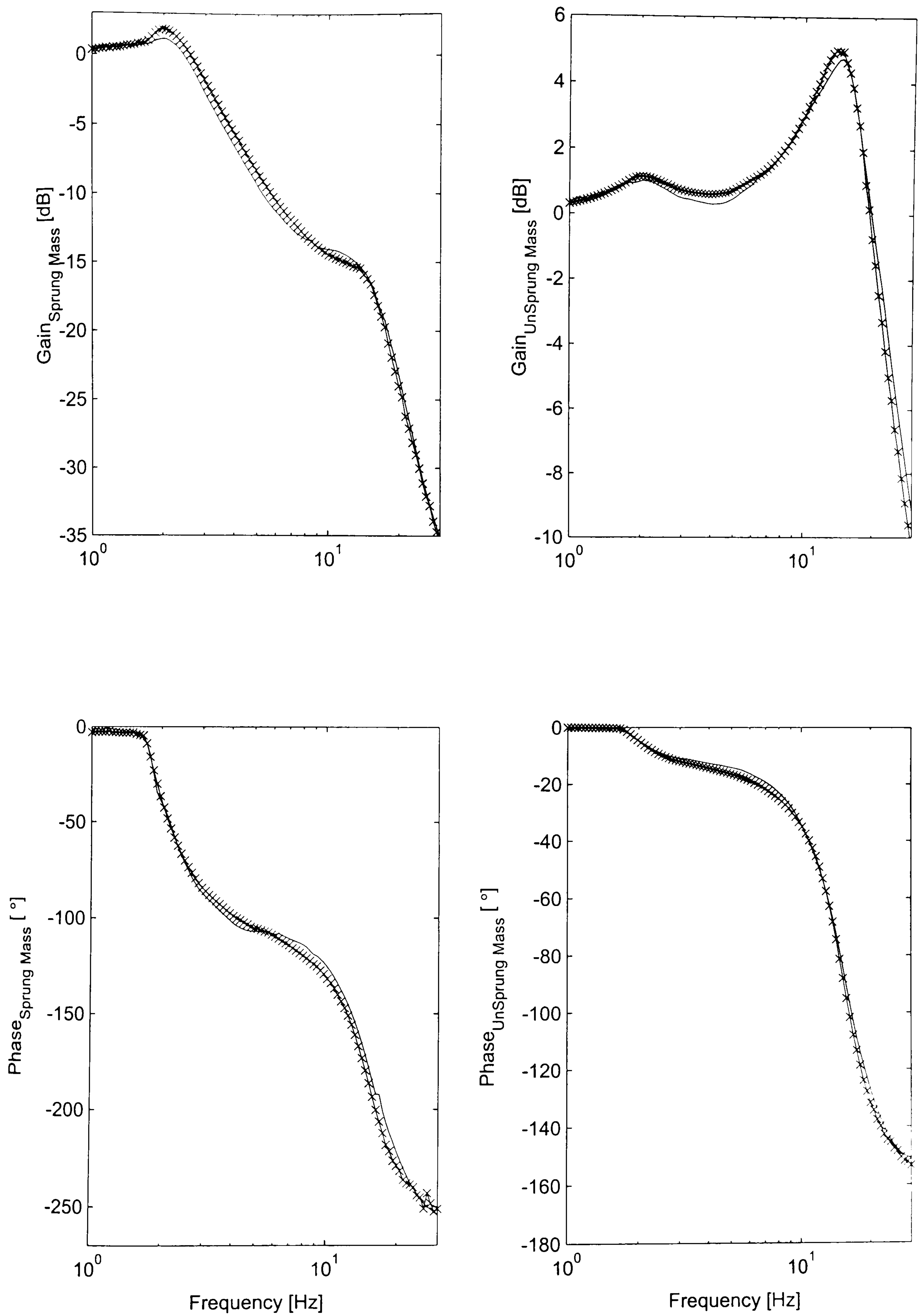


Fig. 4.5.13 Plots of fitting result with $\text{RMS}_{\text{input}} = 7.5 \text{ mm}$; --measured, -x- modelled.

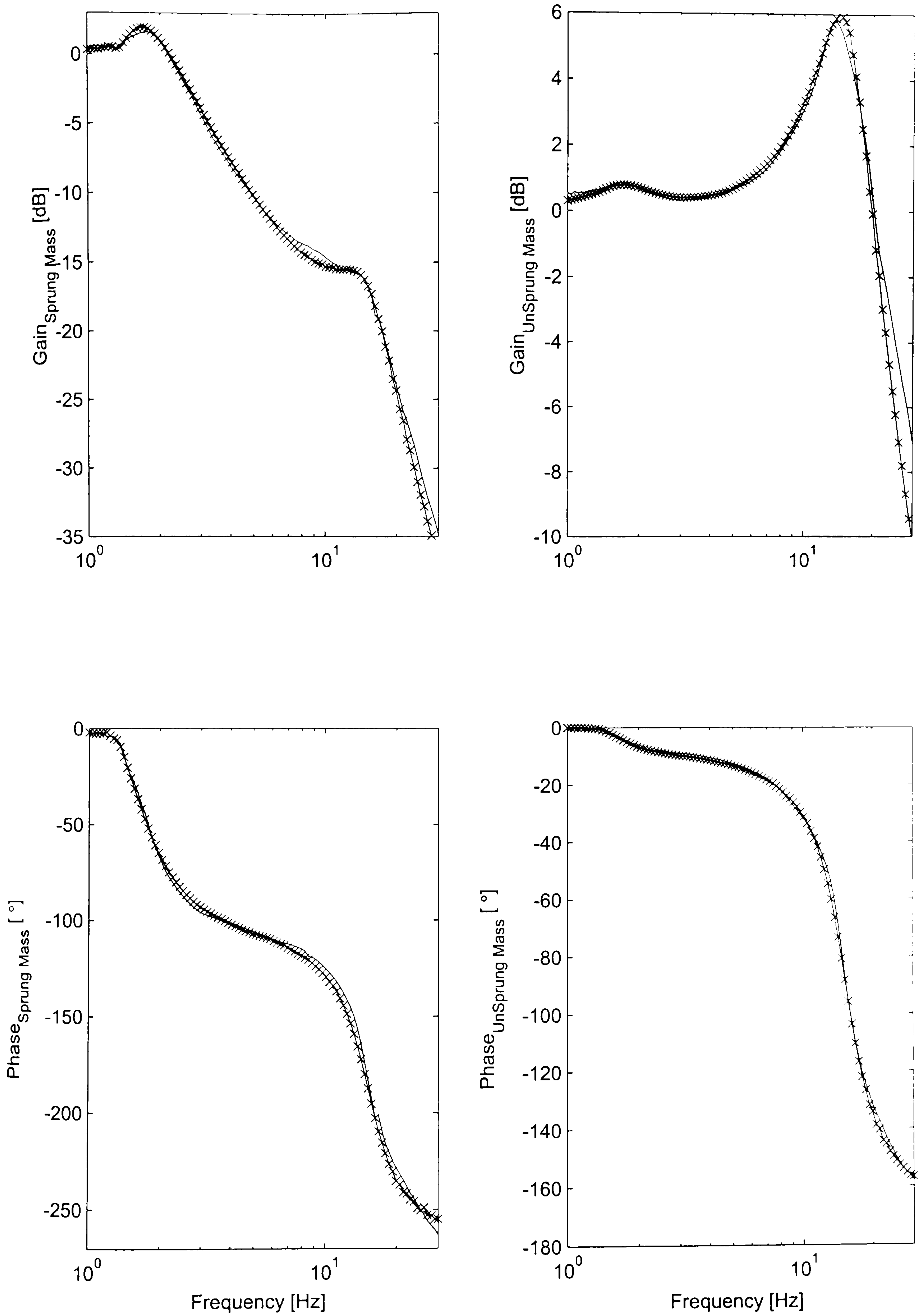


Fig. 4.5.14 Plots of fitting result with $\text{RMS}_{\text{input}} = 12.5$ mm; --measured, -x- modelled.

4.6 Conclusion

In this chapter, the system identification and parameter estimation was applied to a single wheel station. The investigation was a development toward a more complicated problem for parameter estimation of a full vehicle, which will be presented in the next chapter. The identification and parameter estimation was carried out in both the time and frequency domains. The analyses on linear and nonlinear models were discussed. The resultant model that described the dynamic of the single wheel station with an acceptable accuracy was Model 3. Model 3 comprised of sprung and unsprung masses, with suspension and tyre spring, which were modelled as linear spring elements, and the damper was modelled as a spring in series with a bilinear dashpot. The model's quality of fit of the model measured in term of MSE was approximately 7%. With the use of more information on system and measurement noise properties, the model quality of fit was improved with the highest MSE value being below 4%. As far as modelling is concern the modelling error mainly arise from nonlinearities in the spring-damper sub-assembly (strut). Apart from the nonlinear behaviour of the damper, which was amplitude, frequency, and temperature dependent, the strut assembly tended to introduce further nonlinearities in the form of friction and hysteresis. A more accurate model may be obtained with an expense of more modelling effort required and an increase of complexity in the numerical solution and parameter estimation procedure. The performance of the parameter search methods were discussed and compared as complexity of minimisation problem increased. The GB method was shown to suffer from local minima problems. The DS method encountered local minima many times and had a slow convergence rate. The DE and

dvHDE methods were able to find the optimum point. The DE several times reached a situation where the mutation and crossover operations could not produce a better offspring that won over its corresponding parent. This resulted in the DE failed to find the optimum point. When converging, the DE had a steady convergence rate. However, there were several consecutive generations that the DE method failed to update the best chromosome. The dvHDE has shown to be the most favourable method amongst the methods considered. Measured in term of the time required to reach the optimum point, the dvHDE required 19% less computational time, and was shown to consistently find the optimum point.

CHAPTER 5 PARAMETER ESTIMATION OF A LANDROVER 110

5.1 Introduction

The identification and parameter estimation technique developed in the previous chapters is applied to a full wheeled vehicle in this chapter. The investigation is performed on a Landrover 110, using a four-post suspension test facility. The tasks to be carried out are the same as in the case of the ‘single wheel station’ of the last chapter; that is to choose the best model among a set of models considered and to estimate the model parameters. However, the system identification and parameter estimation task is more difficult due to the following two main reasons. First, the dynamic behaviour of the vehicle is complex because of the coupling together of the different modes [3,4] and the nonlinear characteristics of the components involved [1,2]. The second is that it involves a larger number of parameters to be identified. To successfully obtain a vehicle model and its parameters, the problem must be divided into several stages. Due to time limitation, the results presented in this chapter are from the initial investigation when the interest has been focused on the vehicle dynamics in the bounce and pitch modes. A model considered in the identification and parameter estimation process was derived from a four degrees of freedom (4DOF) model. An investigation into the problem that considers the vehicle in bounce, pitch as well as roll motions and employs a more complex model has been left for future work. Some preliminary results from the identification using a 4DOF model that takes into account the effect of the vehicle suspension linkages, and a seven degrees of

freedom (7DOF) model, which allows for vehicle body bounce, pitch and roll and the axle roll, are provided in Appendix C.

This chapter is organised as follows. Section 5.2 describes the vehicle test facilities, the four-post suspension test rig, and the experimental set-up. Various tests were carried out enabling both the time history and frequency responses to be collected and discussed. Equations of motion were used to generate a model for the vehicle under investigation. The model was a planar model (2-dimension in longitudinal and vertical planes) having four degrees of freedom, and it is described in Section 5.3. The estimation of the model parameters was formulated as a multi-objective optimisation problem in the frequency domain. Having in mind that parameter estimation in the time and frequency domain are complementary, and both have their advantages and disadvantages, however due to limited amount of time, the parameter estimation in the time domain was left for future work. When estimating in the frequency domain, the objectives are to minimise the differences between the frequency response of the vehicle model and the experimentally obtained frequency response. The frequency response is a standard test and commonly used for experimental dynamics investigations. The parameter estimation in the frequency domain thus has an advantage of ease of access to experimental data. Use of accelerometers allow the vehicle response at various locations to be measured conveniently. The estimation procedure is described in section 5.4 and the results are presented in section 5.5. Finally, the conclusions were drawn in section 5.6 regarding the model quality and the performance of different numerical search methods.

5.2 Vehicle Testing

5.2.1 The Four Post Test Facility

The vehicle used in this study was a L110 Long Wheelbase Military specification Landrover, see a photograph in Fig. 5.2.1. The vehicle has four-wheel-drive and consists of two beam axles. The rear axle is located longitudinally by trailing arms, and laterally by an 'A' frame upper link mounted to the vehicle chassis. Coil springs and inclined dampers are used, see the bottom left photograph of Fig. 5.2.1. Similarly the front axle is located longitudinally by leading arms. Coil springs and dampers are used, the dampers are installed vertically through the coil springs, see the bottom right photograph of Fig. 5.2.1. Lateral control of the front axle is provided by using a Panhard rod.

Experiments involved in this study were performed at the Engineering Dynamic Centre, Royal Military College of science, Cranfield University, UK. The vehicle tests were performed on the four post hydraulic rig, also shown in Fig. 5.2.1. The test rig consisted of four independently driven, high bandwidth electro-hydraulic actuators; these had a maximum dynamic stroke of 150 mm. The test rig is mounted rigidly to a steel base, which in turn is fixed to the floor. The test rig is controlled using displacement feedback by a Type 3142 servo control unit manufactured by Servo Consultants Ltd. The control unit has the ability of controlling each of the four actuators separately by the use of four separate voltage signal inputs. Fig.5.2.2 provides a layout of the main equipment of the four post test rig. The wheels of the vehicle were placed on top of rectangular plates, called 'wheel pans', that in turn are attached to the top of the actuators, thus enabling the wheels to be excited by

simulated ground inputs. A ground input from a computer or arbitrary signal generators can be sent to the controller unit, which then controls the motion of the actuators.



Fig.5.2.1 Four Poster Test Rig and a Landrover 110 front and rear suspension.

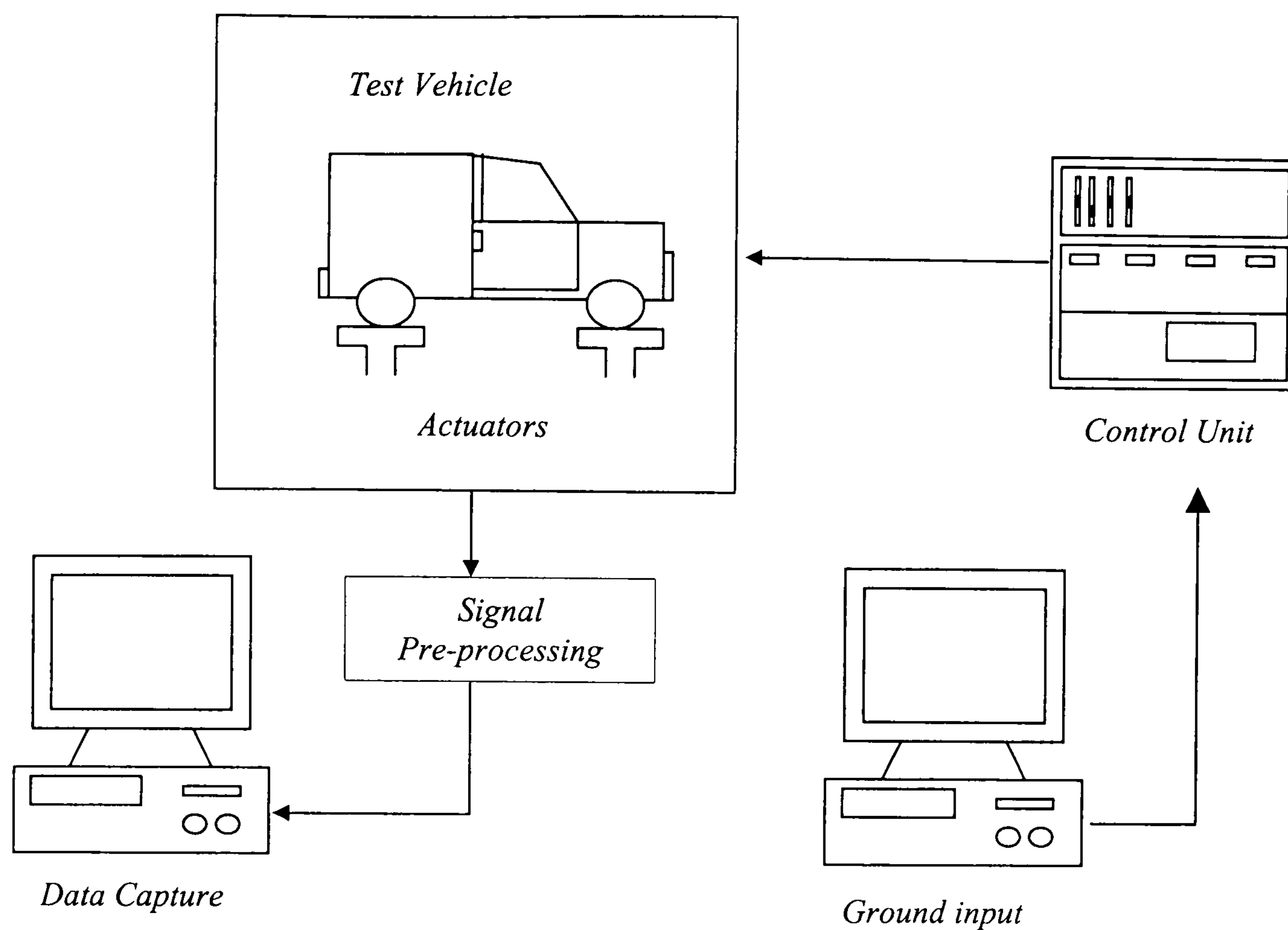


Fig. 5.2.2 A layout of the Four post test facility

The vehicle responses were measured using a variety of sensors. The dynamic and static tyre contact forces were measured using force transducers, also called load cells, installed in the four wheelpans. The actual ground input to the vehicle, that is the wheelpan displacements, was measured using the rig's internal transducers, permanently installed within each actuator housing. The displacements such as that of the vehicle body and wheel-axle masses were measured by using the available potentiometers and linear voltage displacement transducers, LVDTs, attached to appropriate positions of the test vehicle. The required acceleration measurements were obtained by using accelerometers attached to the vehicle at various positions. The required measurements were filtered using lowpass filters with a cut-off

frequency of 30 Hz. The sensor signals were then collected by data capture software that was capable of collection 16 signals at the same time using another computer.

5.2.2 Time Response

The tests carried out on the vehicle consisted of both time response and frequency response tests. During the initial investigation of the vehicle, its response to different types of input were examined. These included sine waves at different frequencies and amplitudes, step inputs and swept sine where the input signal was continuously increasing in frequency from 0.5 to 30 Hz and decreasing in amplitude. Fig. 5.2.3 and Fig. 5.2.4 shows typical plots of the time history of the vehicle suspension working space measured across the suspension springs, and the dynamic tyre force at the four road wheels, when the vehicle was excited with all four actuators driven in phase by the same swept sine input.

The plots of suspension working space, which are the relative displacements of the vehicle body to the axles, show the nonlinear dynamic behaviour of the vehicle suspension in the frequency range of interest. First, it can be seen in Fig. 5.2.3 that there are two peaks. These belong to the vehicle body bounce and wheel hop resonant frequencies. At the beginning of the excitation and up to about 10 seconds in the plots, i.e. low frequencies interval, it can be seen that the suspension working space at the front, both the right and left hand sides, are approximately the same amplitude in positive and negative directions. This suggests that the suspension at the front

behaved in an approximately linear fashion. At about the peak, the amplitude of the left side was slightly higher than that of the right side.

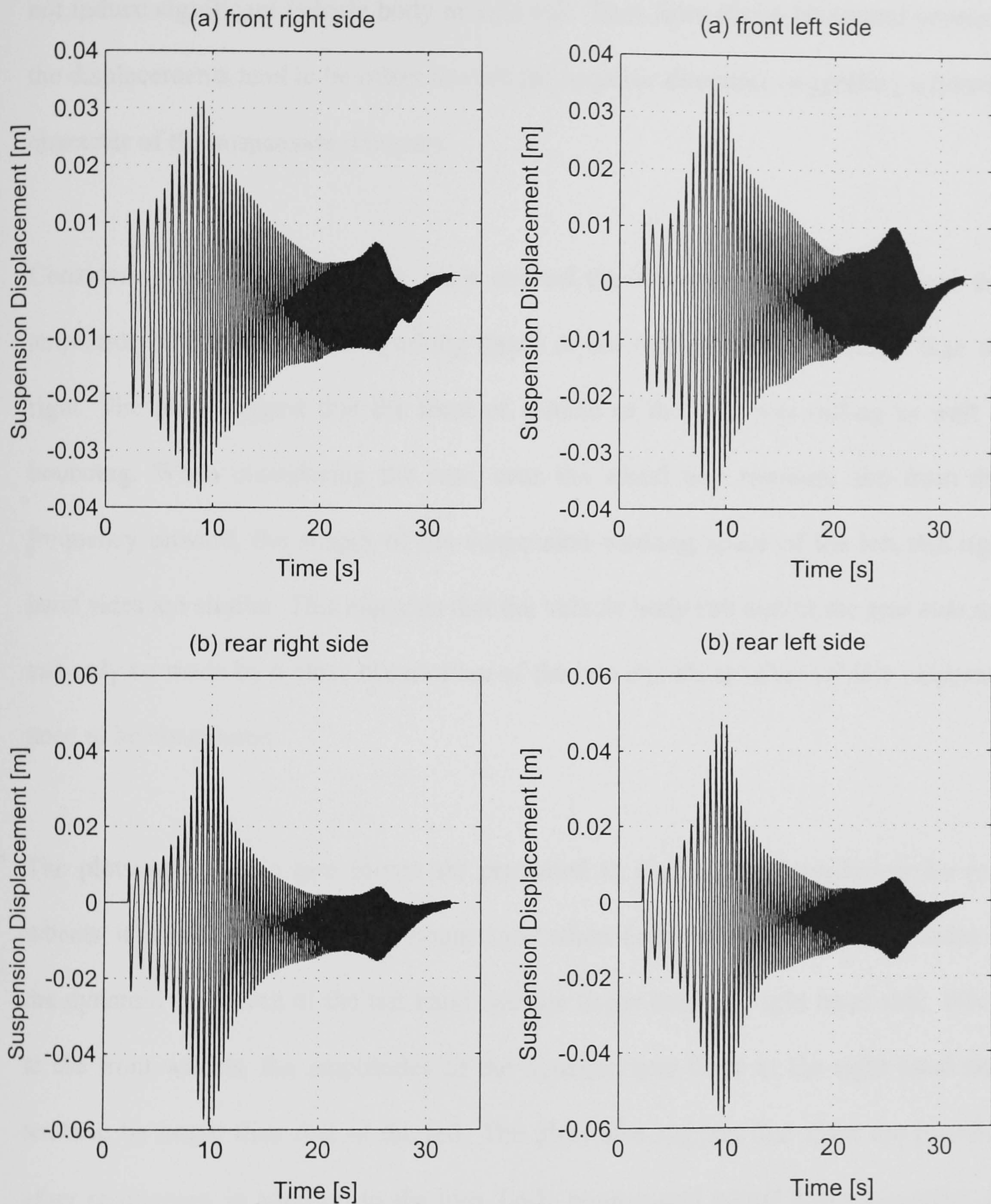


Fig. 5.2.3 Time history of vehicle suspension working space; (a) front, (b) rear.

Considering the same interval for the rear, it can be seen that the suspension working space in the positive and negative directions are similar in amplitude, and the plot of

the rear left side is very similar plot of the rear right side. This suggest that the vehicle suspension at the rear corners behaved approximately linearly and the excitation did not induce significant vehicle body or axle roll. Then from about 10 second onwards, the displacements tend to be offset toward the negative direction, suggesting a bilinear character of the suspension dampers.

Considering the front, when the input excited the wheel hop mode, it is seen that amplitude of the suspension working space at the left hand side is larger than the right. The plots suggest that the front of vehicle or the axle was rolling as well as bouncing. When considering the rear, near the wheel hop resonant and from that frequency onward, the shapes of the suspension working space of the left and right hand sides are similar. This suggests that the vehicle body roll and/or the rear axle roll can only be made by a close observation of the two signals or other vehicle responses need to be considered.

The plots of dynamic tyre forces are presented in Fig. 5.2.4. Considering the rear wheels, it is seen that at the body bounce and wheel hop resonances, the amplitudes of the dynamic tyre force of the left hand side are larger than the right hand side. While at the front wheels, the amplitudes of the dynamic tyre force at the right hand side tends to be larger than that of the left. The plots also suggest that there are possibly other resonances in addition to the two, body bounce and wheel hop resonances, as can be seen clearly on the plot of the front right hand side wheel. The first additional peak occurred slightly after the excitation has passed the body bounce mode may be due to the vehicle body pitching motion. The second additional peak occurred after

the wheel hop mode possibly caused by the engine motion, which is mounted at the front end of the vehicle.

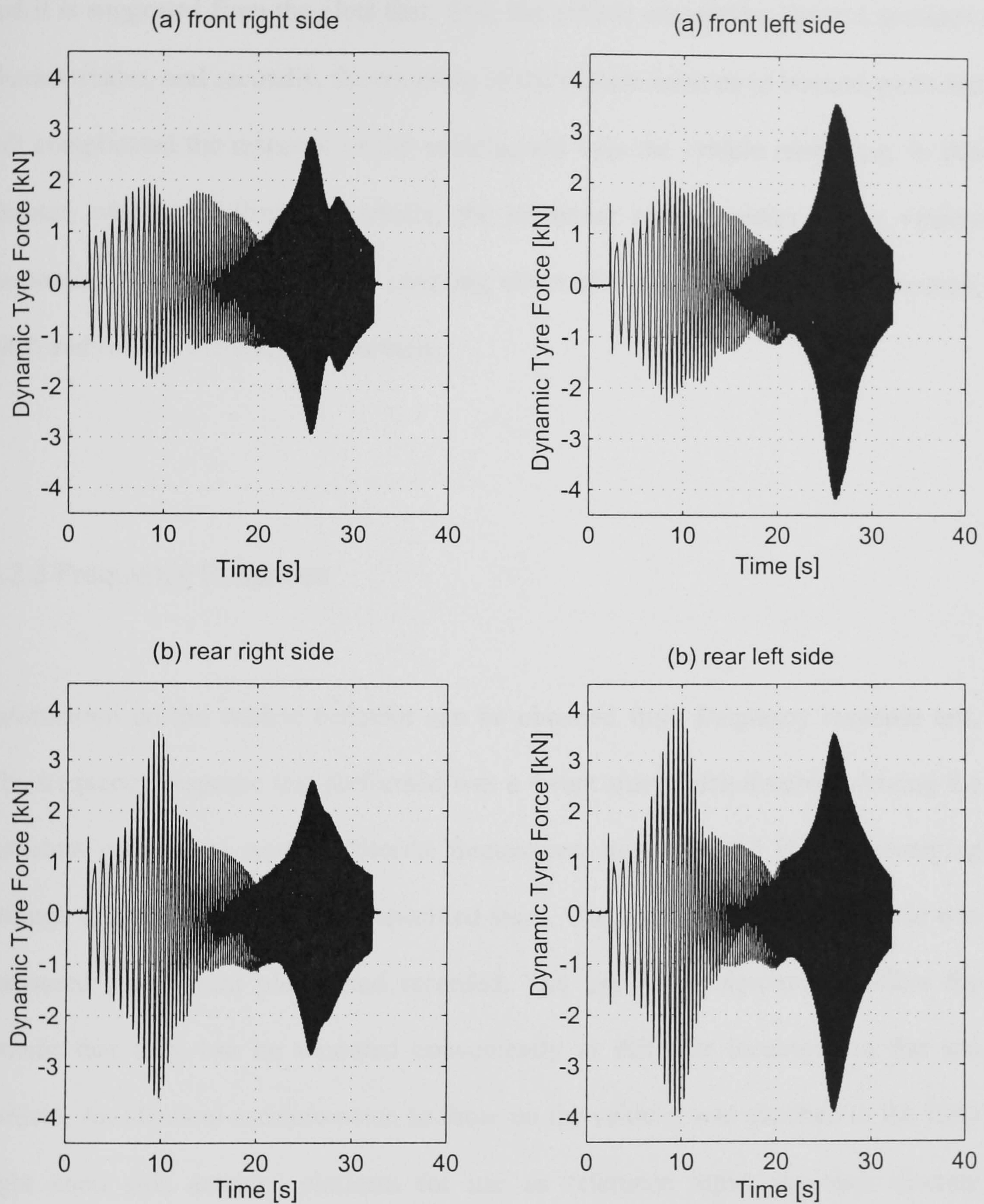


Fig. 5.2.4 Time history of dynamic tyre forces; (a) front wheels, (b) rear wheels.

The analysis of the time history plots of the vehicle responses, such as presented here, indicate that the suspension working space and the dynamic tyre force signals give

useful information for the modelling of the vehicle model for identification and parameter estimation purposes. Here the vehicle motion in bounce was considered, and it is suggested from the plots that, first, the vehicle suspension showed nonlinear characteristics, and secondly, the coupling of the vehicle motions in bounce, pitch and roll complicated the response of the vehicle, and thus the vehicle modelling. In this chapter, when modelling the vehicle, the nonlinear characteristics of the vehicle suspension is simplified, and the coupling effect of the vehicle motions in bounce, pitch and roll are considered separately.

5.2.3 Frequency Response

Information on the vehicle behavior can be obtained from frequency response test. The frequency response test performed was a swept sine which involved driving the actuators with a sine wave at discrete frequencies, starting at 0.5 Hz and sweeping through to 30 Hz in a number of specified steps. The acceleration of the vehicle was measured at different places and recorded. The use of accelerometers offers the benefit that they can be mounted conveniently at different locations on the test vehicle. An identical accelerometer, to those on the vehicle, was attached to the front right hand side actuator platform for use as reference input. At each discrete frequency, the accelerations of the vehicle and actuator were measured over a number of sine wave cycles. The amplitudes of the measured accelerations at the discrete excitation frequencies were calculated automatically by the computer control equipment, thus enabling the amplitude of the vehicle accelerations to be compared to the amplitude of that of the reference ground input. This method of testing is the

standard transfer function test. There were initially five accelerometers attached to the vehicle. One accelerometer was mounted on the front left hand side wheel, another one was on the back left hand side wheel. These two accelerometers on the wheels would thus measure the response of the front and rear unsprung masses respectively. One accelerometer was mounted at the center of gravity of the vehicle and would measure vehicle bounce motion. One accelerometer was at the front of the vehicle on the centre line of the vehicle (mid-track). This would be used to measure the vehicle pitch motion. An accelerometer mounted to the left of the vehicle in line with the centre of gravity would be used to measure vehicle roll motion. Additional accelerometers could be mounted and used for further investigation as required.

The vehicle frequency response tests were carried out in three phases; the first with the vehicle in bounce, the second with the vehicle excited in pitch, and the third with the vehicle excited in roll. In the bounce transfer function tests, all four actuators were driven simultaneously in phase at the same amplitude and frequency. Similarly, the pitch test was carried out with the rear two actuators 180° out of phase relative to the front two actuators. The actuators movement would then induce primarily vehicle pitch motion. In the vehicle roll test, the right hand side actuators were driven 180° out of phase to the left hand side actuators. The actuator movement would induce primarily vehicle roll motion. The test results are presented and discussed below.

Figures 5.2.5 and 5.2.6, respectively show the frequency response of the vehicle body and axles in the bounce motion test. Four accelerometers were placed at the four wheel-corners of the vehicle chassis, vertically above the axles. Three more accelerometers were mounted at different locations but all were inline with the vehicle

centre line; the first one was at the middle of the front chassis cross member, the second one was at the rear towing point, and the third one was at the vehicle centre of gravity.

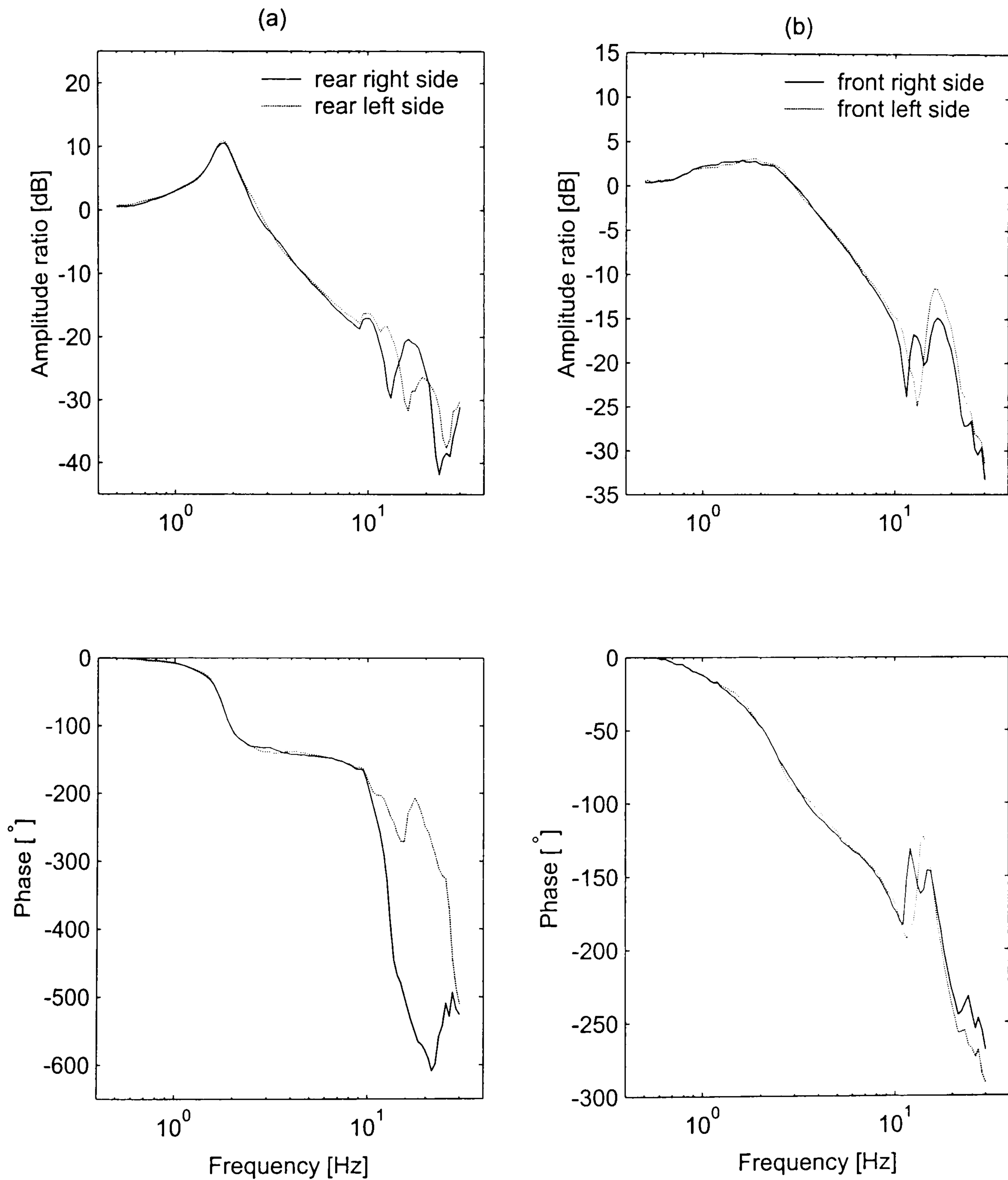


Fig. 5.2.5 Plots of vehicle body frequency response in bounce; (a) rear corners, (b) front corners.

In Fig. 5.2.5, the resonance frequencies of the vehicle rear corners are approximately 1.80 Hz. The amplitude ratios and phases of the rear right hand side and left hand side

are very similar up to about 10 Hz, after which they start to deviate. The resonance peaks for the front corners are less obvious. The maximum amplitude ratio recorded took place at an excitation frequency of approximately 1.7 Hz. The amplitude ratios of the front corners are less than that of the rear corners.

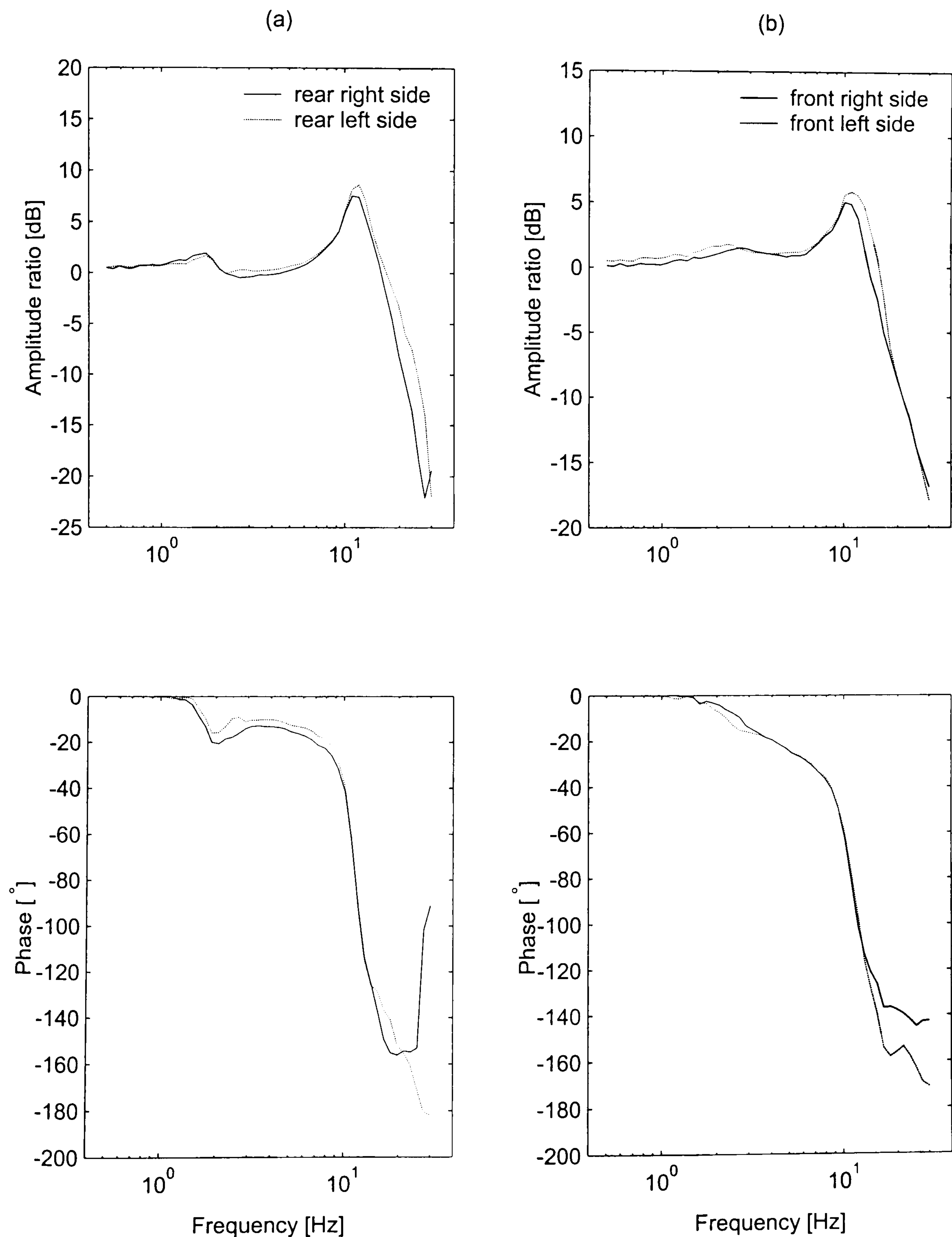


Fig. 5.2.6 Plots of frequency response in bounce; (a) rear axle, (b) front axle.

The wheel hop resonance frequencies for the rear and front axles are approximately 11 Hz and 10.5 Hz respectively, see Fig. 5.2.6. The amplitude ratios and phases of the right and left hand sides are similar but not the same.

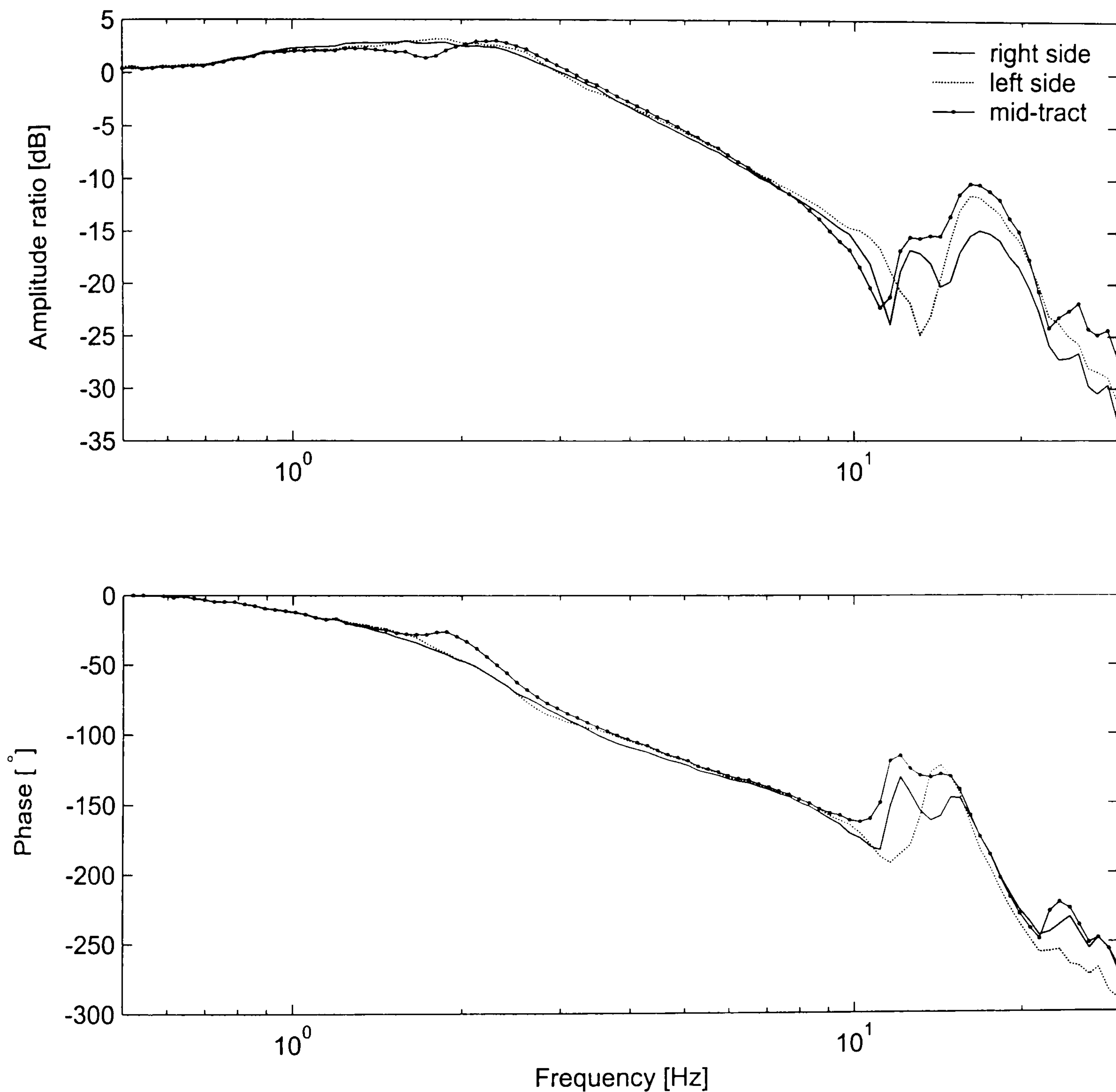


Fig. 5.2.7 Plot of vehicle body frequency response (front end).

It can also be seen from the plot of the vehicle frequency response at the middle of the cross member of the front chassis in Fig. 5.2.7 that there was another peak, which occurred after the bounce mode. This is the pitch resonance of the vehicle body. The

differences between them can be seen from the low frequencies, with larger differences after the wheel hop modes. This was to be expected because of the axle roll and that their centres of mass were a small distance from the vehicle centre line.

5.3 Vehicle Models

In order that the parameter estimation procedure can be carried out, both the measured and modelled responses are required. The responses from a model are usually obtained by performing a simulation on a computer, where the model equations of motion are solved. The mathematical model for the vehicle under test was developed in a number of stages, and at each stage the response of the model to various inputs was compared to that of the real vehicle. This approach would ensure that the model was built in accordance with the available experimental data.

Three models were initially considered, two of them can be derived by hand, the other is generated using a commercial multi-body dynamics software. The first and second both are four degrees of freedom models and contains 12 parameters to be identified. The second model is different from the first model in that it takes into account the effect of the vehicle suspension linkages. The model is thus nonlinear and derivation by hand is not impossible. The third model is a seven degrees of freedom (7DOF) model and contains 22 parameters to be identified.

To assess the quality of each model in describing the dynamic behaviour of the vehicle under test, a search for the ‘best’ parameters according to the cost function set

must be performed, using one of many numerical search techniques available. For a chosen search method, it usually requires a reasonable number of search runs so that the comparison between the models is fair and statistically justified. Similarly, for a given model, a reasonable number of runs need to be carried out in order to compare and then justify the performance of different numerical search methods. Due to a limited amount of time, only a small number of runs have been carried out for the second and third models, both of which take considerably time to complete each parameter estimation cycle. As a consequence, the descriptions and qualities of fit of the second and third models are provided separately in Appendix C. The results of the investigations carried out in the previous chapters have shown that the GB method tends to suffer from local minima problem, the DS method experiences both local minima and slow convergence problem, and the DE method has a slower convergence speed. The parameters of the second and third models were therefore chosen to be identified using the dvHDE method.

This section describes the derivation of the equations of motion of the first model. The model is then used in the parameter estimation procedure and the comparison of four different numerical search methods; the GB, DS, DE and dvHDE methods.

Model 1: vehicle bounce and pitch model

The first stage in the development for a full vehicle model of the Landrover 110 was the creation of a four degrees of freedom (4DOF) model. The model is a planar model in 2-dimensional planes (longitudinal and vertical) as depicted in Fig. 5.3.1, see also

Appendix B for the definition of the vehicle axes. This model will be referred to as Model 1.

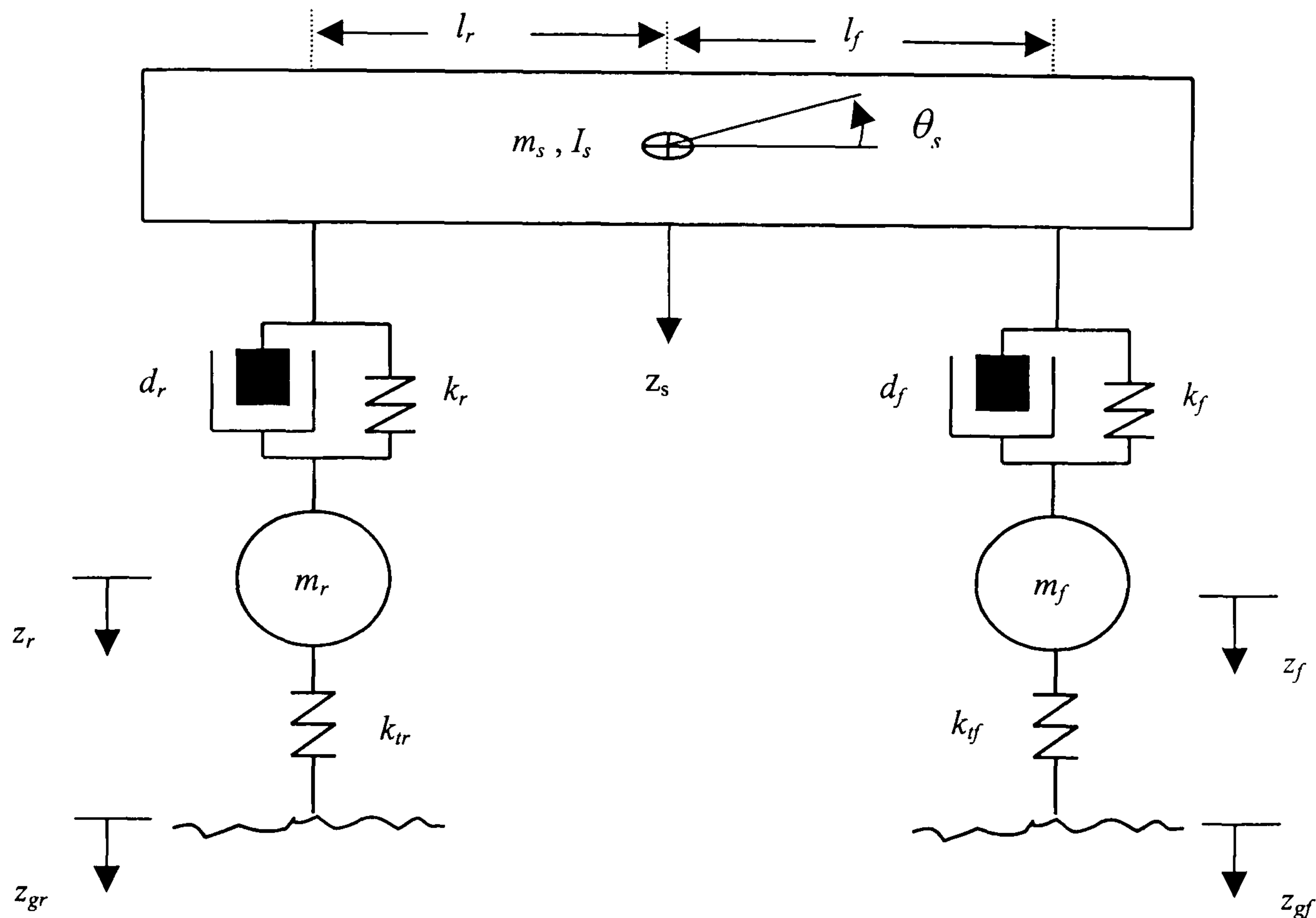


Fig. 5.3.1 A four degrees of freedom (4DOF) vehicle ride model

In Model 1, the vehicle is modelled by three lump masses, m_s , m_f and m_r ; where m_s is the sprung mass representing the vehicle body whose moment of inertia property is denoted by I_s , m_f and m_r are the front and rear unsprung masses representing the front and rear axles and associated hardware, respectively. Noting that, it is difficult to specify precisely which parts of the vehicle should be on the sprung or unsprung masses, therefore m_s , m_f and m_r are the equivalent masses. The compliance and damping properties of the vehicle front and rear suspension springs, dampers and bushes are represented by idealised springs and viscous damping elements, with the

stiffnesses of k_f , k_r and the damping coefficients d_f , d_r for the front and rear, respectively. The sprung mass is supported by the front and rear suspension forces acting at the distance of l_f and l_r respectively away from the centre of gravity. The front and rear tyres are modelled as linear springs elements, where k_{tf} , k_{tr} denoting the tyre stiffnesses of the front and rear types respectively. In addition, it is further assumed that the only point contact occurs at the wheel contact patches.

By applying Newton's second law and using the static equilibrium position as the origin for both the linear displacement of the centre of gravity and angular displacement of the vehicle body, the equations of motion of the model are formulated below. The sign convention is taken to be positive in the direction as shown in Fig. 5.3.1, see also Appendix B.

From the free body diagram provided in Fig. 5.3.2, and assuming small angular displacement of the vehicle body, θ_s , the equations of motion for each mass can be derived as follows.

Consider the sprung mass, the equation of motion in bounce is;

$$m_s \ddot{z}_s = -f_{sr} - f_{dr} - f_{sf} - f_{df} \quad (5.1)$$

The equation of motion in pitch is given by;

$$I_s \ddot{\theta}_s = -l_r (f_{sr} + f_{dr}) + l_f (f_{sf} + f_{df}) \quad (5.2)$$

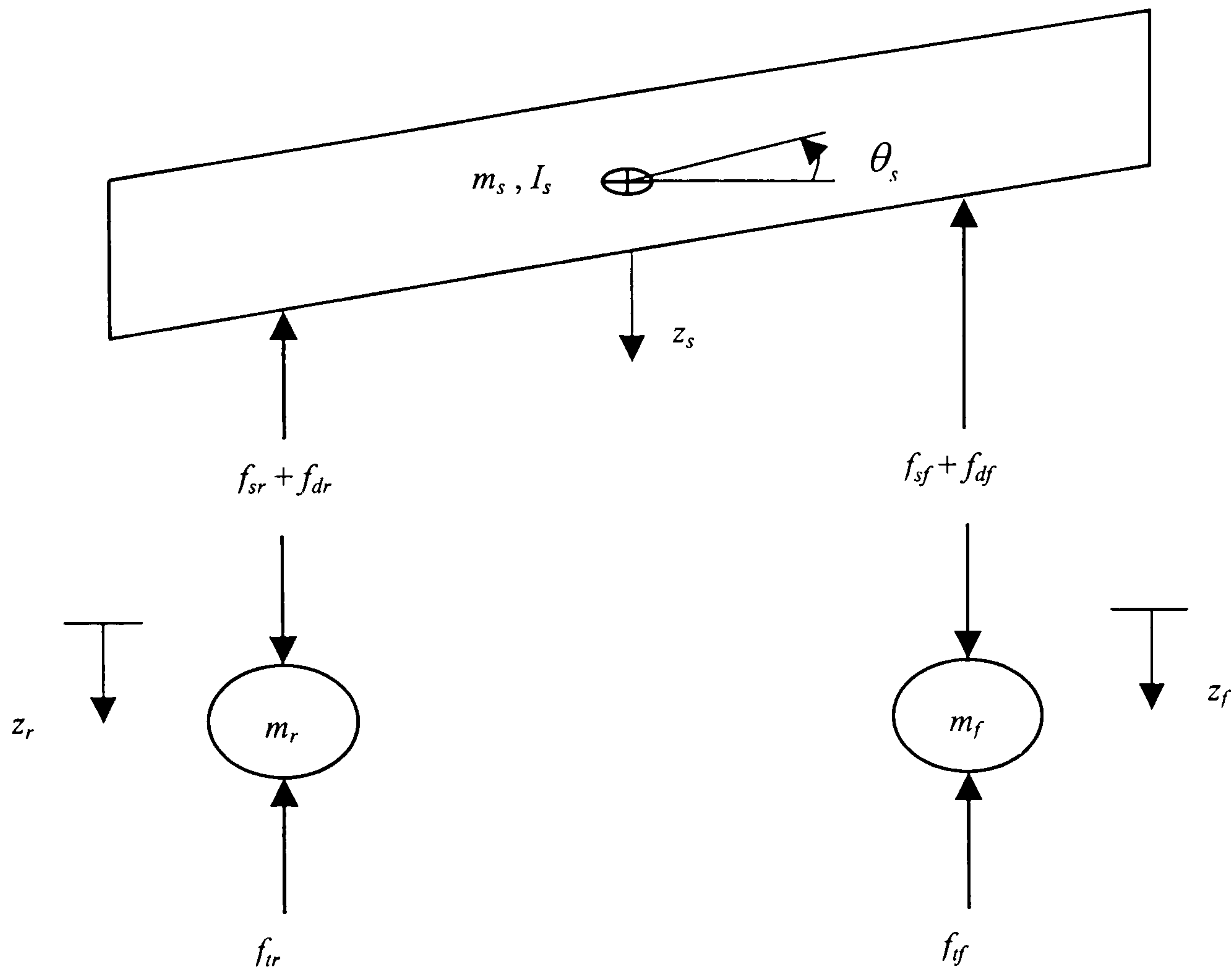


Fig. 5.3.2 Free body diagram

Consider the unsprung masses; the equation of motion of the front unsprung mass is;

$$m_f \ddot{z}_f = f_{sf} + f_{df} - f_{ff} \quad (5.3)$$

Similarly, the equation of motion of the rear axle is;

$$m_r \ddot{z}_r = f_{sr} + f_{dr} - f_{lr} \quad (5.4)$$

where

$$\text{The front suspension spring force, } f_{sf} = k_f (z_s - l_f \theta_s - z_f) \quad (5.5)$$

$$\text{The front suspension damping force, } f_{df} = d_f v_{fsus} \quad (5.6)$$

$$\text{The rear suspension spring force, } f_{sr} = k_r (z_s + l_r \theta_s - z_r) \quad (5.7)$$

$$\text{The rear suspension damping force, } f_{dr} = d_r v_{rsus} \quad (5.8)$$

$$\text{The front tyre force, } f_{tf} = k_{tf} (z_f - z_{gf}) \quad (5.9)$$

$$\text{The rear tyre force, } f_{tr} = k_{tr} (z_r - z_{gr}) \quad (5.10)$$

and

$$d_f = \frac{d_{fr}}{2} (1 + \text{sign}(v_{fsus})) + \frac{d_{fb}}{2} (1 - \text{sign}(v_{fsus})) \quad (5.11)$$

$$d_r = \frac{d_{rr}}{2} (1 + \text{sign}(v_{rsus})) + \frac{d_{rb}}{2} (1 - \text{sign}(v_{rsus})) \quad (5.12)$$

$$v_{fsus} = (\dot{z}_s - l_f \dot{\theta}_s - \dot{z}_f) \quad (5.13)$$

$$v_{rsus} = (\dot{z}_s + l_r \dot{\theta}_s - \dot{z}_r) \quad (5.14)$$

where d_{fb} and d_{fr} are the damping coefficients in the compression and extension directions belonging to the front suspension, respectively. Similarly for the rear suspension, d_{rb} and d_{rr} are the damping coefficients in the compression and extension directions, respectively. The function *sign* returns 1 if v_{fsus} or v_{rsus} is positive and returns 0 if otherwise. This means, bilinear damping characteristics are included in the model in order to take into account the different damping forces generated when the vehicle suspension is undergone motions in the bump and rebound directions. In addition, the following notations are used;

$\theta_s, \dot{\theta}_s, \ddot{\theta}_s$ pitch angular displacement, velocity and acceleration of the sprung mass

$z_s, \dot{z}_s, \ddot{z}_s$ the vertical displacement, velocity and acceleration of the sprung mass

z_f, \dot{z}_f the vertical displacements and velocity of the front unsprung mass

z_f, z_r the vertical displacements of the front and rear unsprung masses

z_{gf}, z_{gr} the vertical displacements the front and rear ground inputs

Thus, in model 1, the vehicle is modelled as a planar (2-dimentional) model with linear suspension spring and bilinear damping elements, and the tyres are initially modelled as simple linear springs. The model therefore contains 12 parameters to be identified.

The equations of motion derived above describe the motion of the vehicle in terms of differential equations. The next step is to solve the equations to derive the nature of the response of the system. With the use of Matlab and Simulink, computer codes takes advantage of the computers ability to perform many calculations very quickly to solve these differential equations within a specified error limit. The calculation is performed in such a way that each iteration will be complete before the next is started. Starting with the initial forces acting on the vehicle, the computer can calculate the pitch and bounce accelerations. The computer then calculates the velocities and displacements of the sprung mass and unsprung masses by means of integrating the equations. These quantities are fed back for the next iteration and are used to generate the next set of solutions.

5.4 Parameter Estimation

Having obtained a model, the ‘best’ model parameters need to be identified in order to evaluate the model quality in describing the dynamic behaviour of the vehicle under test. The parameter estimation was carried out in the frequency domain, the estimation in the time domain being left for future work. The technique used is depicted in Fig. 5.4.1.

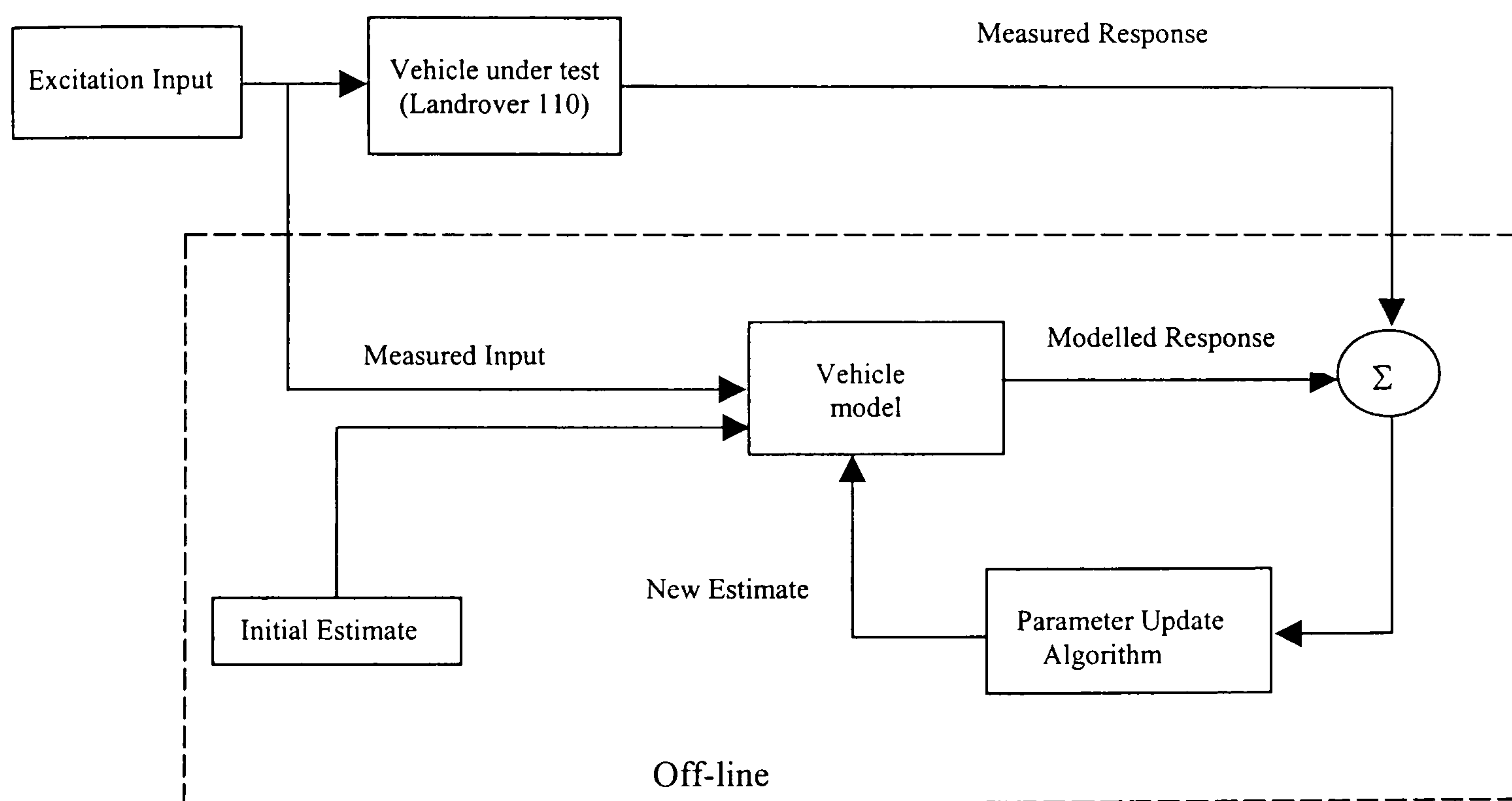


Fig. 5.4.1 Parameter estimation method

The estimation is an off-line procedure as indicated by the dotted-box in Fig. 5.4.1. From the vehicle frequency response test, the gain and phase of the vehicle outputs at different locations relative to the reference input were collected. For an assumed vehicle model, the input and an initial guess of the model parameters are used to produce the model frequency responses. Since the model considered are nonlinear, the

technique developed in section 4.2.3 is thus employed. The technique is derived in the same way as that implemented on the Four Post suspension test rig.

The parameter estimation is a multi-objective optimisation problem in the sense that the differences in both the gains and phases of the sprung and unsprung masses at various locations are to be minimised simultaneously. The minimisation is formulated with the cost function being a weight least squared sum of the difference between the modelled and measured frequency response at different frequency points. The determination for the weighting matrix is not an easy task and is usually carried out by trial-and-error. In this work, the weighting for each frequency points were the variance of the experimentally data (gains and phases) obtained after a few repetition of the frequency response test. Both the gains and phases were used in formulating the cost function. The gains and phases of the sprung and unsprung masses were initially given the same weighting.

The cost function is then in turn used by a numerical search method to update the initial guess of the model parameters. With the new parameters, the process is repeated until the termination criterion set in the algorithm is reached.

5.5 Estimation Results

A typical plot of fitting the vehicle frequency response using Model 1 with the bilinear damper characteristics is shown in Fig. 5.5.1. Despite the model simplicity,

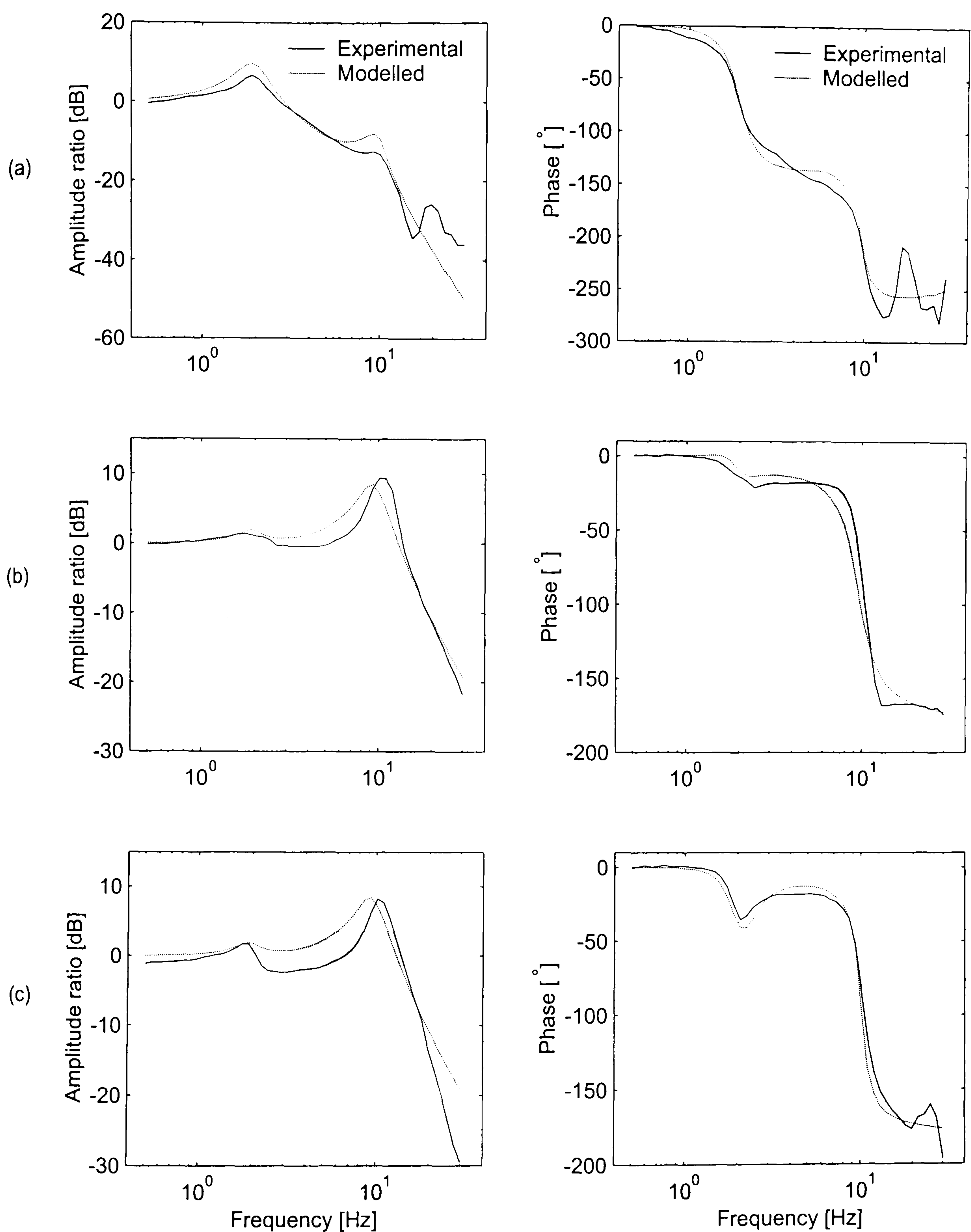


Fig. 5.5.1 Fitting result of vehicle frequency response by Model 1; (a) vehicle body, (b) front axle, (c) rear axle.

it can be said that, the 4DOF model capture the dominant dynamics of the vehicle in bounce excitation relatively well. Considering Fig. 5.5.1 (a), small differences between the modelled and experimental gains (amplitude ratio) and phases are seen at low frequencies, below 1 Hz, and the differences show to increase as frequency increases toward the vehicle body bounce resonance. Larger difference can be observed around the body bounce and wheel hop modes, when compared to the intermediate frequencies (between the two resonance modes) and low frequencies. After approximately 15 Hz, where the modelled responses do not follow the experimental responses, the differences in both the gain and phase can be clearly seen toward the ends of the plots. This suggests that the model was capable of describing the dynamic behaviour of the vehicle body up to the wheel hop mode, after which it failed. However, large difference seen toward the ends of the plots, where the resonance of the engine occurred, are to be expected since the effect of the engine behaviour was not included in the model.

Considering the front axle in Fig. 5.5.1 (b), again, it can be said that the model was capable of capturing the dynamic behaviour of the axle, though the differences between the modelled and experimental responses can be seen. Larger differences between the experimental and modelled gains and phases are seen at the intermediate frequencies (between the body bounce and wheel hop modes) when compared to low frequencies (below the body bounce mode) and high frequencies interval (approximately from 10 Hz to 15 Hz). After approximately 15 Hz, the differences are shown to increase again as frequency increases. The differences toward the end are

influenced by the effect of the engine resonance and the axle roll. The same arguments can be made when considering the rear axle.

The performance of Model 1 in describing the vehicle under test is thus justified that it is capable of capturing the dominant dynamic behaviour of the vehicle under test up to the wheel hop mode. The largest difference between the modelled and experimental response is observed above the frequency of approximately 15 Hz, where the engine resonance and the front and rear axle roll become significant.

Apart from the model incapability to describe the vehicle behaviour at high frequencies, the differences between the experimentally obtained and modelled gains and phases in the frequency interval of 0.5-15 Hz comes from several sources. Addition to the errors caused by the techniques of obtaining the gains and phases (both the technique implemented on the Four-Post test rig and that developed in section 4.2.3), the main source of the error is due to the assumptions and simplicity of the model. The estimated model parameters and thus the quality of the model are influenced by the choice of the cost function formulated. The weightings given to the gains and phases for each discrete frequency point, which were then used to form the cost function, were all unity. The ‘improper’ choice of these weightings may also contributes to the total difference between the experimental and modelled responses.

A more realistic 4DOF model that takes into account the effect of suspension linkages and a more complex model that allows for the vehicle body bounce, pitch and roll as well as the front and rear axle roll are described in Appendix C. Due to time limited, only some preliminary estimation results are provided there.

Performance of different numerical search methods

The parameters of Model 1 were estimated using different numerical search methods; the GB, DS, DE and dvHDE methods. The performances of the four methods considered are presented in Table 5.4.1. The results from 20 independent runs are compared in terms of the mean and standard deviation of the cost function, m_J and σ_J , number of function evaluations, $m_{T_{CPU}}$ and $\sigma_{T_{CPU}}$, and the CPU-time required, $m_{T_{CPU}}$ and $\sigma_{N_{feval}}$. The time required to complete one independent run for each method was measured in hours. It was mainly spent in the determination of the model frequency responses and searching for the best 12 parameters.

	m_J	$m_{T_{CPU}}$ [hr.]	$m_{N_{feval}}$	σ_J	$\sigma_{T_{CPU}}$ [hr.]	$\sigma_{N_{feval}}$
GB	1070.10	2.58	1317	501.2	0.21	131
DS	870.53	3.18	3184	163.4	0.13	251
DE	751.64	2.01	2216	125.1	0.27	285
dvHDE	750.45	1.75	1519	132.5	0.23	173

Table 5.4.1 Performance of different numerical search methods

From Table 5.4.1, the results from the GB method show high values of both mean and standard deviation values of the cost function. This indicates that the method encountered the problem of obtaining local minima, thus spent more time, when compared to the DE and dvHDE methods, searching in the wrong places. The DS

method was able to find the optimum point, however it was trapped in local minima several times. When it converged, it has shown to have a slow convergence rate and took a longer time to reach the same value of the cost function when compared to the DE and dvHDE methods. As a result, the mean value of the cost function for the DS method is high, and the CPU-time is the highest among the four methods.

The DE and dvHDE methods were able to find the optimum point. The final cost function values, m_f , are similar. The value of the DE was slightly higher than that of the dvHDE method due to a small number of runs that the DE terminated as the maximum generation set was reached while the values of the cost function was still slightly high. The dvHDE algorithm control parameters were set for a fast convergence speed, while monitoring the population diversity and number of function evaluation so that to prevent premature due to local minima and mis-convergence due to stagnation. The control parameters were set as follows; the population size, $n_i = 20$, the mutation factor, $\mu_m = 0.7$, the crossover factor, $\mu_c = 0.8$. The Acceleration operation performed 5 iterations, and it was carried out only when the reduction of the best fitness value of the current generation from that of the previous generation was less than 10 percent. The Migration was activated when the population diversity was below 0.1. The required precision for the model parameters, θ_{tol} , were set as follows; the masses ± 0.1 kg, the stiffness ± 0.01 kN/m, the damping coefficients ± 1 Ns/m. From the result given in Table 5.4.1, the dvHDE method has shown to require a less number of function evaluations, and spent less time to reach the same quality of fit when compared with that of the DE method, measuring in term of CPU-time the dvHDE performed better by approximately 13 %. The dvHDE thus performed better over the GB, DS and DE methods for the problem being considered.

5.6 Conclusion

The identification and parameter estimation of a Landrover 110 has been divided into several stages. Due to a limited amount of time, the initial investigation into the problem has been focused on the vehicle dynamics in bounce and pitch modes, and the parameter estimation has been limited to the frequency domain only.

In this chapter, a vehicle bouncing and pitching four degrees of freedom (4DOF) model has been considered for the identification of the vehicle under test. The model equations of motion have been derived and its parameters have been estimated by fitting the frequency response of the model to the experimentally obtained frequency response. Both the gains and phases of the vehicle body, and front and rear axles have been used and given the same significance (weighting) when formulating the cost function in the parameter estimation process. The result has shown that, despite the model simplicity, it was capable of capturing the dominant dynamic behaviour of the vehicle under test in the frequency range of interest up to the wheel hop mode (approximately 10 Hz), after which the difference between the experimental and modelled response was shown to deviate. The performance of the model in describing the vehicle's dynamic behaviour has been as expected, though large difference between the modelled and experimental responses was observed at high frequencies, above approximately 15 Hz, where the resonance of the engine took place and its effect together with the axle roll became significant. Apart from the model's inability to capture the dynamics at frequency beyond the wheel hop mode, the differences between the experimental and modelled frequency response are mainly due to the

assumptions and simplicity of the model and possibly the effect of ‘improper’ choice of the weighting given to the data at each discrete frequency point.

Bearing in mind the limitation of the model and the parameter estimation carried out in the frequency domain, the best model parameters were obtained using four difference numerical methods. The GB and DS methods experienced problems of premature termination due to local minima and had slow convergence speed. The DE was shown to take 13% longer time to complete one cycle of the parameter estimation procedure when compared to the dvHDE method. For the model being considered and with the parameter estimation carried out in the frequency domain, the dvHDE was therefore shown to be the best method among the methods considered.

CHAPTER 6 SUMMARY, CONCLUSIONS and FUTURE WORK

6.1 Summary

In this thesis, an optimisation method, named the discrete variable Hybrid Differential Evolution (dvHDE) method, has been developed and proposed to solve the parameter estimation problem. The development has aimed to overcome problems such as premature termination due to local minima, mis-convergence and time consumption due to slow convergence speed, which were experienced by the conventional and global optimisation methods. The ultimate aim of developing the dvHDE method was therefore to improve convergence speed without sacrificing the consistency of finding the global optimum solution.

The dvHDE method is a population-based method that does not require calculation of derivatives of the objective function. Its working mechanism relies on the recombination and natural selection based on the principle of survival of the fittest. In the recombination (mutation and crossover operations), the parent (current) population is used to produce a population of offspring. An offspring then has a one-to-one competition with its corresponding parent in the selection for the population of the next generation. The dvHDE method is based on the Differential Evolution (DE) algorithm of Storn and Price, however it is fundamentally different from the original DE in two important aspects.

- First, the dvHDE method employs integer-encoding technique, and deals with parameter estimation problems containing different types of parameters in the same unified manner. In the dvHDE algorithm, the parameters, which can either be integer, discrete or continuous variables or a combination of them, are all treated as discrete variables. When the parameter involved is an integer, it is directly treated as a discrete variable, and if the parameter is a continuous variable, it is discretised using available information about its required precision and lower and upper search bounds. Having employed an integer-encoding technique and treated all parameters as discrete variables, the dvHDE augments an extra gene to the chromosome. This extra gene is the identifier of the chromosome in the entire search space, and its purpose is for checking and avoiding re-evaluating of the chromosomes that have already been assigned a fitness values.
- The second difference is that the dvHDE algorithm includes two additional mechanisms, the acceleration and migration operations. The acceleration operation can be used to improve the convergence speed. The migration operation can be used to maintain and/or improve the population diversity, thus reducing the possibilities of obtaining local minima as experienced by the gradient-based (GB) and Downhill Simplex (DS) methods, or problem of stagnation experienced by the DE method. The acceleration operation is basically a local search routine such as the steepest descent or simplex methods. In this work, the Downhill Simplex (DS) of Nelder and Mead has been employed in the acceleration operation. The acceleration operation is optional and is performed only a small number of times in order to maintain the

population diversity. The migration is also optional and is performed only when a given criterion is met. In the operation, a newly generated population is injected to the current population in an attempt to escape from the present region of the search space.

In this thesis, the dvHDE method has been applied to a number of vehicle related problems of increasing complexity. These include, first, three test problems, which have been studied by several researchers, and later, three practical case studies involving the parameter estimation from experimental data of automotive dampers, a single wheel station (a simplified system representing a quarter of a vehicle), and a full vehicle. The estimation results are as summarised sections 6.1.1 and 6.1.2.

6.1.1 Test problems

The performance of the dvHDE method has been demonstrated and validated against a number of other techniques through the investigation on three test problems. The investigation on the test problems provided an opportunity to explain how the dvHDE algorithm could be setup and explicitly illustrate the problems that might be encountered during the search for the best parameters, problems such as premature termination due to local minima and mis-convergence due to stagnation, and how the dvHDE overcame them.

A single degree of freedom (1DOF) mass-spring-damper

The first test problem was a simulated single degree of freedom (1DOF) mass-spring-damper problem, where the parameter involved were all continuous variables. The DE and dvHDE methods tend to be slower than the GB method. This is generally true when the search space contains no local minima and the correct model structure is assumed. In such a case, the GB method has an advantage of efficiently making use of the information about the gradient of the objective function. Nevertheless, the simulation study has shown some advantages of the dvHDE method over the original DE methods. Firstly, the dvHDE required less function evaluations to reach the same optimum point. Though the difference in the number of function evaluation was low, 3%, it must be remembered that the dvHDE method had also included the migration, a mean of preventing premature termination problem. Secondly, the dvHDE was shown to reduce the ‘discontinuous’ convergence characteristic experienced by the DE method, and thus the dvHDE reached the optimum point in 43 generations compared to 83 generations of the DE method.

A gear train design

The second test problem was the optimisation of a gear train design, in which the parameter involved were all integers. The dvHDE has shown to be able identify all of the four optimum points. The results were compared to eight other methods including three gradient-based methods (Sequential Quadratic Programming, Nonlinear Programming,

and Sequential Linearisation Algorithm), two Genetic Algorithms (Modified Genetic Algorithm and Meta-Genetic Algorithm), Simulated Annealing, Evolutionary Programming, and Differential Evolution (DE) methods. The dvHDE has found an equivalent result to that of the GAs and DE methods and a better result than the other five. The investigation identified and explicitly explained two distinct convergence problems, premature termination due to local minima and mis-convergence due to stagnation, and how the dvHDE overcame them using the migration and acceleration operations. The consistency of achieving the global optimum point was shown to increase with a larger population size, however, at the expense of a longer computational time. The introduction of the acceleration and migration in combination with increase of the population size was shown to improve the percentage of successful identification from 62% to 98% and reduced the computational time required by 36%.

A coil spring design

The third test problem was the optimisation of a coil spring design involving a mixture of integer, discrete and continuous variable. The estimation result was compared to four other techniques including a gradient-based method (Sequential Quadratic Programming), Genetic Algorithm, Meta-Genetic Algorithm, and the DE method. The DE and dvHDE methods were able to find the same result that was better than the gradient-based method and the two genetic algorithms. With the incorporation of the acceleration and migration operations, the dvHDE method had the highest success rate of 99%, and when compared with the original DE method, the dvHDE required up to 80%

less number of function evaluations. The performance of the dvHDE method therefore was shown to be the best compared to the other techniques listed.

6.1.2 Practical case studies

The performance of the dvHDE method has been further investigated, with the estimation of parameters from experimental data for three practical case studies relating to the dynamics of wheeled vehicle component, subassembly and a full vehicle. The followings are the summaries of each case. The conclusions are made with regarded to both the identification aspect as well as the performance of the dvHDE method compared with other techniques applied to the same problems.

Parameter estimation of an automotive damper

The first case study was the parameter estimation of an automotive damper over the frequency range of interest was 0.5-30 Hz. In this investigation, five damper models of increasing complexity were proposed in the selection of the best model for the damper under test. For each model, four different numerical search methods (the GB, DS, DE and dvHDE methods) were applied to estimate the model parameters, and their performances were compared. As far as the selection of the best model among the set considered is concerned, the model, which consisted of a linear spring connected in series with a bilinear viscous damping element, was shown to be the best one, in the sense that it gave

the best compromise between model quality of fit and the computational time required to complete the estimation procedure. Among different numerical search methods, the results from 50 independent runs have shown that the GB method had a low success rate and suffered with problem of obtaining local minima. The DS method had a low convergence, and was occasionally trapped in local minima. The DE and dvHDE methods have shown to be able to find the optimum point every run. However between the two methods, the dvHDE has shown to perform better by almost 30% in term of CPU-time.

Parameter estimation of a Single Wheel Station

The second case study was the parameter estimation of the ‘single wheel station’, a simplified system representing a quarter of a wheeled vehicle. The investigation has been carried out to test the identification and parameter estimation technique, which will later be applied to a full vehicle. The parameters of the single wheel station, which represents a quarter of a medium-sized family car, have been successfully identified both in the time and frequency domains.

The parameter estimation in the time domain was formulated as a multi-objective optimisation using five system output signals. A set of models was considered, and their qualities in describing the dynamic behaviour of system being studied in the frequency range of, 0.5-30 Hz, were discussed. The best model consisted of sprung and unsprung masses, with the suspension and tyre springs modelled as linear spring elements and the

damper as a linear spring connected in series with a bilinear damping element. This model gave the best quality of fit in term of the normalised mean squared error (MSE) of 7%, and when incorporating information about the system and measurement noise properties via a Kalman Filter (KF) gave an MSE of 4%. The discrepancies between the modelled and measured system outputs are thought to mainly arise from nonlinearities in the spring-damper subassembly (strut).

Four different numerical search methods have been applied to estimate the parameters of the single wheel station. The results from the four methods were discussed and compared as the complexity of the parameter estimation problem increased, both due to the numerical calculation and the number of parameters to be identified. The gradient-based (GB) method was shown to experience problems of obtaining local minima, while the Downhill Simplex (DS) method encountered a combination of two problems, obtaining local minima and a slow convergence speed. The Differential Evolution (DE) method was shown to experienced problems of stagnation and ‘discontinuous’ descent characteristic that these have resulted in the DE failing to find the optimum point or to otherwise converge at a slow rate. The dvHDE method was shown to perform better over the other methods considered, as it did not experience local minima and stagnation problems, thus it consistently found the optimum point, and the convergence speed was improved, compared to the DE method, by 19% measured in term of CPU-time.

The parameter estimation in the frequency domain was formulated as a minimisation of the weighted sum of the difference between the experimental and modelled frequency

responses. The experimental data were obtained from the system frequency response test, and for the model, a method based on the Fourier series was developed for calculating the frequency response of a nonlinear model. The results have shown that the model captures the dynamic behaviour of the system being studied well. The difference between the measured and modelled frequency responses are thought to come mainly from the limitation of the model and the errors in the parameter estimation procedure when formulating the objective function. An additional error was due to the errors from the method of obtaining the experimental and modelled frequency response.

Parameter estimation of a Landrover 110

Finally, the identification and parameter estimation procedure was applied to a Landrover 110. Due to time limitation, only the results from the initial investigation, which focused on the vehicle bounce and pitch dynamic behaviour, were presented. The investigation considered a four degrees of freedom (4DOF) model in the identification process, and the model parameter estimation was formulated as a multi-objective minimisation in the frequency domain, where both the gains (amplitude ratio) and phases of the system outputs relative to the input were used in the objective function.

The estimation results showed that the model was capable of describing the dominant dynamic behaviour of the vehicle under test well up to the wheel hop mode, after which the effect of the engine resonance and the front and rear axle roll caused the difference between the modelled and measured frequency responses. Apart from the model

incapability in capturing the dynamic behaviour at high frequencies, above approximately 15 Hz onward, the difference between the modelled and measured frequency responses are thought to arise mainly from the simplicity of the model. Additional errors were due to the errors in the parameter estimation procedure when assigning the weighting in the cost function and in the methods of obtaining the frequency responses implemented on the Four-Post test facility and that developed for nonlinear models.

Bearing in mind the limitation of the model and the parameter estimation, which has only been carried out in the frequency domain, the 4DOF model was used to compare the performance of four different numerical search methods. The GB method was shown to experience problem of obtaining local minima and the DS method was shown to have a slow convergence speed. The DE was also shown to take a long time to complete one cycle of the parameter estimation procedure when compared with the dvHDE method. The dvHDE method was shown to require less number of function evaluations, and spent less time to reach the same quality of fit when compared with that of the DE method, measuring in term of CPU-time the dvHDE method performed better by approximately 13%. Thus for the problem considered, the dvHDE was shown to be the best method among the methods considered.

6.1.3 The performance of the dvHDE method

In this thesis, the newly developed discrete variable Hybrid Differential Evolution (dvHDE) method has been applied to a limited number of vehicle related problems of increasing complexity. The parameter estimation problems studied so far are small and medium sized where the number of parameters to be identified is up to twelve parameters. The results have shown that, apart from simple problem, which is when the optimisation is one or two dimensional, the dvHDE method has shown to be the best method for all problems investigated. The method has shown to perform better than the other techniques with varying percentages of computational time and success rate. As problem complexity increased the dvHDE method has been shown to have advantages over the conventional (the gradient-based and simplex methods) and global optimisation methods (the Differential Evolution method). This has been achieved mainly from the combination of the main working mechanism (mutation and crossover operations) and the two additional operations (the acceleration and migration operations), and some possible benefit from checking and avoiding re-evaluation of the repeated solutions, this however depends on the problem size and the characteristics of the objective function.

The dvHDE method is a population-based method, thus it is not susceptible to problem of obtaining local minima when the objective function is non-differentiable, nonlinear or multi-modal. The acceleration operation helps speed up the convergence and the migration operation can be performed if experiencing premature termination and/or mis-convergence problems. The dvHDE's performance (the ability and consistency to find the global optimum point) is influenced by a small set of the algorithm control parameters

(population size, mutation and crossover factors) and also the strategies for the acceleration and migration operations. The set-up for both the algorithm control parameters and the acceleration and migration operations should always reflect the best compromise between convergence speed and consistency of obtaining the global optimum point. Initially the dvHDE method should be attempted without the acceleration and migration operations. When the search shows signs of encountering premature termination and mis-convergence problems, the dvHDE with migration operation should be attempted first. A combination of the acceleration and migration operations can then be included if the problems of premature and/or mis-convergence is encountered, and convergence speed is of important.

6.2 Conclusions

- A new numerical optimisation method, named the discrete variable Hybrid Differential Evolution (dvHDE) method, has been developed and shown to solve a number of parameter estimation problems. The method is a population-based technique and its working mechanisms mimic biological process that evolves using recombination and natural selection based on the principle of survival of the fittest. The proposed method has two distinctive features; first, it employs an integer-encoding technique and treats all parameter involved in the same manner as discrete variables, secondly, it embeds two mechanisms that have been shown to overcome

convergence difficulties and speed up the iterative search in the optimisation procedure.

- The new method has been validated against a number of other techniques (including gradient-based methods, Genetic Algorithms, Simulated Annealing, Evolutionary Programming, and Differential Evolution) using three test problems. The test cases were nonlinear programming optimisation problems relating to mechanical and engineering design applications, and involved different types of variables (continuous, integer and a mixture of integer, discrete and continuous variables). The results have shown the dvHDE method to be both efficient and effective for this type of problem.
- The dvHDE method has been successfully applied to small and medium sized problems, where the number of parameter to be estimated is less than thirteen. Apart from optimisation problems in one or two dimensions, the proposed dvHDE method has been shown to have an increasing advantage over the other methods considered. The technique is thus applicable to nonlinear dynamics problems of low order, in particular where the objective function in the optimisation is commercially expensive.
- The performance of the proposed dvHDE method depends on its control parameters. A general guidance for setting the control parameters has been given, however their values are problem dependent, usually the set-up should reflect the best compromise between convergence speed and the consistency of finding the global optimum point.

The success of the dvHDE method thus depends on the careful selection of the control parameters.

- The developed technique has been applied to number of experimental problems where its robustness has been demonstrated. The method was shown to successfully identify the parameters for a range of models of increasing complexity from experimental data. This makes a possible method to be used for other engineering application.
- In the particular areas of vehicle suspensions a working tool has been presented for parameter estimation. The technique has been shown to successfully estimate the parameters for suspension components, a single wheel station [127] and a full vehicle in both the time and frequency domains.

6.3 Future Works

- Possible benefits of the discrete variable Hybrid Differential Evolution (dvHDE) method need exploring further, overcoming the limited number of investigation in this thesis with different types and size of problems.
- More complex vehicle problems, especially the pitch and bounce model, and use of the experimental data into the model to improve the model behaviour. This then can be used to select the best damper/spring combination.

- Applications of the dvHDE method to other problems, not just the vehicle problem but possibly to other applications such as control gain selection of a missile or aircraft control.
- Since in the dvHDE method, the evaluation of the objective function is performed in parallel fashion. A possibility of reducing time in the parameter searching procedure is to develop a strategy to carry out the objective function evaluation using a network of computers.

REFERENCES

- [1] **Purdy, D.J.** Theoretical and experimental investigation into an adjustable automotive damper, Proc. Instn. Mech. Engrs, Part D, Journal of Automobile Engineering, 2000, 214(D3), pp. 265-283.
- [2] **Surace, C., Worden, K. and Tomlinson, G. R.** An improved nonlinear model for an automotive shock absorber, 1992, Nonlinear Dynamics, Vol. 3, pp. 413-429.
- [3] **Gillespie, T. D.** Fundamentals of Vehicle Dynamics, Society of Automotive Engineering, Inc, 1992.
- [4] **Sharp, R. S. and Pilbeam, C.** Achievability and value of passive suspension design for minimum pitch response. In Proceedings, of International Conference on Vehicle Ride and Handling, 15-17 November 1993 (Institution of Mechanical Engineers).
- [5] **Worden, K.** Data processing and experiment design for the restoring force surface method, Part 1: Integration and differentiation of measured data, Mechanical Systems and Signal Processing, 1990, Vol. 4 part 4, pp. 295-319.
- [6] **Eykhoff, P.,** System Identification – Parameter and State Estimation, John Wiley and Sons, New York, NY, 1974.
- [7] **Ljung, L.** System Identification: Theory for the user. Prentice Hall, Inc, 1999.
- [8] **Shoukens, J. and Pintelon, R.** *Identification of linear systems: A practical guideline to accurate modelling.* Pergamon Press. 1991.
- [9] **Walter, E. and Pronzto, L.** Identification of Parametric Models from Experimental Data, Masson, 1997.
- [10] **Sodertrom, T. and Stoica, P.** System Identification, Prentice-Hall, Hertfordshire, UK, 1988.
- [11] **Smith J. M.** Mathematical Modeling and Digital Simulation for Engineering and Scientists, John Wiley and Sons, New York, 1977.
- [12] **Greenwood, D. T.** Principles of Dynamics, Prentice-Hall, Inc., 1988.
- [13] **Moon, F. C.** Applied Dynamics: with applications to Multibody and Mechatronic Systems, John Wiley and Sons, Inc., 1998.
- [14] **Baruh, H.** Analytical Dynamics, The McGraw-Hill Companies, Inc., 1999.
- [15] **Rao, S. S.** Engineering Optimization: Theory and Practice, Wiley Inter-Science, New York, 1996.

- [16] **Goodwin, G. C. and Payne, R.L.** Dynamic System Identification: Experiment Design and Data Analysis, Academic Press, Inc, 1977.
- [17] **Isermann, R.** Practical aspects of process identification, Automatica, 1980, Vol.16, pp. 575-587.
- [18] **Leontaritis, J. L. and Billings, A. S.** Experimental design and identifiability for non-linear systems, International Journal of Systems Science, 1987, Vol.18, No. 1, pp. 189-202.
- [19] **Beck, J. V. and Arnold, K. L.** Parameter Estimation in Engineering and Science, John Wiley and Sons, New York, NY, 1977.
- [20] **Goodwin, G. C. and Sin, K. S.** Adaptive Filtering, Prediction and Control, Printice-Hall, Englewood Cliffs, NJ, 1984.
- [21] **Haykin, S.** Adaptive Filter Theory, Prentice-Hall, Englewood Cliffs, NJ, 1986.
- [22] **Johnson, C., R. Jr.** Lectures on Adaptive Parameter Estimation, Prentice-Hall, Englewood Cliffs, NJ, 1988.
- [23] **Ljung, L. and Soderstrom, T.** Theory and Practice of Recursive Identification, MIT Press, Cambridge, MA, 1987.
- [24] **Treichler, J. R., Johnson, C. R. and Larimore, M. G.** Theory and Design Adaptive Filters, John Wiley and Sons, New York, NY, 1987.
- [25] **Hajdansinski, A. K. Eykoff, P., Damen, A. A. H. and van den Boom, A. J. W.** The choice and use of different model sets for system identification, IFAC Identification and system parameter estimation, Washington D.C.,1982, pp. 47-55.
- [26] **Eykoff, P.** On the coherence among the multitude of system identification methods, IFAC Identification and system parameter estimation, Washington D.C., 1982, pp. 31-42.
- [27] **Mottershead, J. E. and Friswell, M. I.** Model updating in structural dynamics: A survey, Journal of Sound and Vibration, 1993, No.2, Vol. 167, pp. 347-375.
- [28] **Burden, R. L. and Faire, J. D.** Numerical Analysis, 3rd ed., PWS Publishers, Boston, MA, 1985.
- [29] **Fries, R. H. and Cooperrider, N. K.,** Bayesian Estimation of Transit Rail Vehicle Parameters, Journal of Dynamic Systems, Measurement and Control, Transactions of ASME, 1985, Vol. 107, No. 2, pp 151-158.
- [30] **Norton, J. P.** An Introduction to Identification, 1986, Academic Press, London.
- [31] **Astrom, K. J. and T. Bohlin,** Numerical Identification of Linear Dynamic

- Systems for Normal Operating Records, Proc. 2nd IFAC Symposium on the Theory of Self-Adaptive Control Systems, NPL Teddington, England, Plenum Press, New York, 1965, pp. 96-111.
- [32] **Astrom, K. J.** Maximum Likelihood and Prediction Error Methods, *Automatica*, 1980, Vol. 16, pp. 551-574.
 - [33] **Grace, A.** Optimization Tool Box, for use with Matlab. The MathWorks Inc., 1994.
 - [34] **Zalzala, A. and Flemming, P.** editors. Genetic Algorithms in Engineering Systems, ISBN 0 85296 902 3. The Institution of Electrical and Mechanical Engineers, London, UK, 1997.
 - [35] **Thithi, I.** Control system parameter identification using the population based incremental learning, In UKAC International Conference on Control, IEE 1996, pp. 1309-1314.
 - [36] **Lin, Y. and Kortum, W.** Identification of System Physical Parameters for Vehicle Systems with Nonlinear Components, *Vehicle System Dynamics*, 1991, Vol. 20 Supplement, pp. 354-365.
 - [37] **Hall, R. L, Triangle, T. L., and Klinger, D. L.,** Application and Evaluation of System Identification Techniques to Rail Vehicle Dynamics, Systems Control Incorporated Technical Report TR 5307-100, Nov. 1979.
 - [38] **Gelb, A.** Applied Optimal Estimation, 1974, The Analytic Sciences Corporation.
 - [39] **Kalman, R. E. and Bucy, R.** New results in Linear Filtering and Prediction, *Journal of Basic Engineering (ASME)*, Vol. 83D, 1961, pp. 95-108.
 - [40] **Lewis, F. L.** Optimal Estimation with Introduction to Stochastic Control Theory, 1986, John Wiley & Sons, Inc.
 - [41] **Satoru, O. and Aichi-ken, T.** EP 0 808 733 A2, European Patent Application, 1997.
 - [42] **Paul, J. Th. V. and Karl, N.** Vehicle Dynamics Estimation Using Kalman Filters, *Vehicle Dynamics*, 1999, Vol. 32, pp. 171-184.
 - [43] **Baguley, S. R.** ON the practical implementation and convergence of the extended Kalman filter and other approximate non-linear estimation methods, *International Conference on Control*, 1996, Vol.1, pp 364-368.
 - [44] **a-Velo, J. G. and Walker, B. K.** Aircraft Parameter Identification Using Extended Kalman Filtering, *MATLAB Conference*, 1993

- [45] AGARD Lecture Series No. 104, 'Parameter Identification', 1979.
- [46] **Bolzern, P., Cheli, F., Falciola, G. and Resta, F.** Estimation of the Non-linear Suspension Tyre Cornering Forces from Experimental Road Test Data, *Vehicle System Dynamics*, 1999, Vol. 31, pp. 23- 34.
- [47] **Bendat, J. S. and Piersol, A. G.** Engineering Applications of Correlation and Spectral Analysis, New York, Wiley and Sons, 1980.
- [48] **Wellstead, P. E.** Non-parametric Methods of System Identification, *Automatica*, Vol. 17, 1981, pp. 55-69.
- [49] **Jitendra, K. T.** Identification of Multivariable Stochastic Linear Systems via Spectral Analysis given Time-Domain Data, Proceedings of the American Control Conference, Albuquerque, New Mexico, AACC, 1997.
- [50] **Majjad, R.** Estimation of suspension parameters, Proceedings of the 1997 IEEE International Conference on Control Applications, Hartford, CT, October 5-7, 1997.
- [51] **Bubhardt, J. and Isermann, R.** Realization of adaptive shock absorbers by estimating physical process coefficients of a vehicle suspension system., ACC/WM1, 1992, pp. 531-535.
- [52] **Kim, C. and Ro, P. I.** Reduced-order modeling and parameter estimation for a quarter-car suspension system, Proceedings of the Institution of Mechanical Engineerings, part D, 1999, Vol. 214, pp. 851-864.
- [53] **Kim, C., Ro, P. I. and Kim H.** Effect of the Suspension structure on equivalent suspension parameters, Proceedings of the Institution of Mechanical Engineerings, part D, 1999, Vol. 213, pp. 457-470.
- [54] **Best, M. C. and Gordon, T J.** A randomized integral error criterion for parametric identification of dynamic models of mechanical systems, Proceedings of the Institution of Mechanical Engineerings, part I, 1999, Vol. 213, pp. 119-134.
- [55] **Yu, F. and Crolla, D. A.** 'An Optimal Self-Tuning Controller for an Active Suspension', The active Control of Vibration, Bath, UK, 5-8 September 1994, pp. 51-65.
- [56] **Weispfenning, T. and Leonhardt, S.** Model-based identification of a vehicle suspension using parameter estimation and neural networks, IFAC 13th Triennial World Congress, San Francisco, USA, 1996, pp. 83-88.
- [57] **Gianone, L., Bokor, J. and Hangos, K. M.** Grey Box identification of vehicle

- dynamics, IFAC System Identification, Copenhagen, Denmark, 1994, pp. 1187-1192.
- [58] **Roether, F. and Muller, P. C.** Instrumental variable identification in vehicle dynamics, Fortschr-Ber, Reihe 8, Nr. 114, VDI, Dusseldorf, 1986.
 - [59] **Zhang, H. Y. and Chen, J.** Identification of physical parameters of ground vehicle using block-pulse function method, International Journal of System and Science, 1990, Vol. 21, No. 4, pp. 631-642.
 - [60] **Krishnaswami, V.** A regularization approach to robust variable structure observer design applied to vehicle parameter and state estimation, Proceedings of the American Control Conference, Philadelphia, Pennsylvania, June 1998, pp. 2258-2262.
 - [61] **Gao, J., Leighton, N. J. and Morgan, C.** A low cost transfer function identification technique for automotive suspensions, Vehicle System Dynamics, Vol. 29, 1998, pp. 261-273.
 - [62] **Wurtenberger, M., Germann, St. and Isermann, R.** Modeling and parameter estimation of nonlinear vehicle dynamics, ASMS, DSC, Vol. 44, Transportation Systems, 1992, pp. 53-63.
 - [63] **Yi, K. and Hedrick K.** Observer-based identification of nonlinear system parameters, Journal of Dynamic Systems, Measurement, and Control, June 1995, Vol. 117, pp. 175-182.
 - [64] **Hemingway, N. G.** Immittance identification: An application to the dynamic modeling of vehicle components, International Journal of Vehicle Design, Vol. 6, No. 1, 1985, pp. 55-71.
 - [65] **Germann, St. and Isermann, R.** Determination of the center of gravity height of a vehicle with parameter estimation, IFAC System Identification, Copenhagen, Denmark, 1994, pp. 563-568.
 - [66] **Michelberger, P., Bokor, J., Keresztes A. and Varlaki, P.** Determination of mass, damping and stiffness matrices using structural and parametric identification of linear vehicle frame models, American Control Conference, San Francisco, 1983.
 - [67] **Alloum, A., Charara, A. and Machkour, H.** Parameters non-linear identification for vehicle's model, Proceedings of the 1997 IEEE International Conference on Control Applications, Hartford, CT, October 5-7, 1997.

- [68] **Letty, L. L.** Parameter Estimation in Analytical Models of Automotive Vehicles and Fault Diagnosis, Proceedings of the American Control Conference, Seattle, Washington, June 1995, pp. 1050-1054.
- [69] **Bellizzi, S. and Bouc R.** Identification of the hysteresis parameters of a nonlinear vehicle suspension under random excitation, Nonlinear Stochastic Dynamic Engineering Systems, IUTAM Symposium Innsbruck/Igls, Austria, June 21-26, 1987, pp. 467-476.
- [70] **Kinkel, J. F. and Thomas, M.** Estimation of vehicle dynamic and static parameters from megnetometer data, Journal of Guidance, Control and Dynamics, Vol. 20, No.1, January-February 1997, pp. 111-116.
- [71] **Serban, R. and Freeman, J. S.** Identification and identifiability of unknown parameters in multibody dynamic systems, Kluwer Academic Publisher, 2000.
- [72] **Butz, T., Stryk, O. V. and Wolter, T.** A parallel optimization scheme for parameter estimation in motor vehicle dynamics, Euro-Par 2000-Parallel Processing. Lecture Notes in Computer Science 1900 (Springer-Verlag, 2000), pp. 829-834.
- [73] **Butz, T., Stryk, O. V., Vogel, M., Wolter, T. and Chucholowski, C.** Parallel parameter estimation in full motor vehicle dynamics, SIAM News, Vol. 33. No. 4.
- [74] **Butz, T., Stryk, O. V., Chucholowski, C., Truskawa, S. and Wolter, T.** and Modeling Techniques and parameter estimation for the simulation of complex vehicle structures, High Performance Scientific and Engineering Computing, Lecture Notes in Computational Science and Engineering 21 (Springer-Verlag, 2002), pp. 333-340.
- [75] **Storn, R. and Price, K.** Differential Evolution-A simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization*, Kluwer Academic Publishers, 1997, Vol. 11, pp. 341-359.
- [76] **Price, K. and Storn, R.** Differential Evolution-A simple evolution strategy for fast optimization. *Dr. Dobb's Journal*, 22(4), April 1997, pp. 18-24 and 78.
- [77] **Storn, R. and Price, K.** Minimizing the real function of the ICEC'96 contest by differential evolution. IEEE conference on Evolutionary computation, Nagoya, 1996, pp. 842-844.
- [78] **Price, K. and Storn, R.** Differential Evolution Homepage, www.icsi.berkeley.edu/~storn/

- [79] **Chang, C. S. and Du, D.** Further improvement of optimization method for mass transit signaling block-layout design using differential evolution. *Proceeding IEEE* Vol. 146, No.5, September 1999, pp. 559-569.
- [80] **Kyprianou, A., Giacomini, J., Worden, K., Heidrich, M. and Bocking, J.** Differential evolution based identification of automotive hydraulic engine mount model parameters. *Proc. Instn. Mech. Engrs, Part D, Journal of Automobile Engineering*, 2000, 214(D3), 249-264.
- [81] **Kyprianou, A. and Worden, K.,** Identification of hysteretic systems using the Differential Evolution algorithm. *Journal of sound and vibration*, 2001, 248(2), pp. 289-314.
- [82] **Lampinen, J. and Zelinka, I.** Mechanical Engineering Design Optimization by Differential Evolution. In: David Corne, Marco Dorigo and Fred Glover (editors), 1999. *New Ideas in Optimization*. McGraw-Hill, London (UK), pp. 127-146.
- [83] **Wang, F. and Jang, H.** Parameter estimation of a bio-reaction model by hybrid differential evolution. *Proceedings of the CEC00, 2000 Congress on Evolutionary Computation*, Vol. 1, pp. 410-417. IEEE, Piscataway, NJ, USA. ISBN 0-7803-6375-2.
- [84] **Chiou, J. and Wang, F.** A hybrid method of differential evolution with application to optimal control problems of a bio-process system. *The 1998 IEEE International Conference on Evolutionary Computation Proceedings*, 1998. IEEE World Congress on Computational Intelligence, 1998, pp. 627-632. IEEE, New York, NY, USA.
- [85] **Pedchote, C.** Optimisation of damper setting, 3rd- year undergraduate project, Engineering Systems Department, Royal Military College of Science, Shrivenham, Swindon, UK, 1998.
- [86] **Pedchote, C.** Parameter Estimation of vehicle suspension, Mphil-PhD transfer report, Engineering Systems Department, Royal Military College of Science, Shrivenham, Swindon, UK, 2000.
- [87] **Barnard, S. T. and Simon, H. D.** Fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems, *Concurrency: Practice and experience*, Vol. 6, pp. 101-117, 1994.
- [88] **Pothen, A., Simon, H. D. and Liou, K.** Partitioning sparse matrices with eigenvectors of graphs, *Society for Industrial and Applied Mathematics J. Matrix*

Anal. Appl. Vol. 11, No. 3, pp. 430-452, July 1990.

- [89] **Nelder, J. A. and Mead, R.** A simplex method for function minimisation, *Comput. J.*, 1965, vol. 7, pp. 308-313.
- [90] **Sandgren, E.** Nonlinear integer and discrete programming in mechanical design optimisation. Transactions of the ASME, Journal of Mechanical Design, 112(2): 223–229, June 1990. ISSN 0738-0666.
- [91] **Fu, J.-F., Fenton, R. G. and Cleghorn, W. L.** A mixed integer-discrete-continuous programming method and its application to engineering design optimisation. Engineering Optimisation, 17(4):263–280, 1991. ISSN 0305-2154.
- [92] **Loh, H. T. and Papalambros, P. Y.** A sequential linearisation approach for solving mixed-discrete nonlinear design optimisation problems. Transactions of the ASME, Journal of Mechanical Design, 113(3):325–334, September 1991.
- [93] **Loh, H. T. and Papalambros, P. Y.** Computational implementation and tests of a sequential linearisation algorithm for mixed-discrete nonlinear design optimisation. Transactions of the ASME, Journal of Mechanical Design, 113(3):335–345, September 1991.
- [94] **Zhang, C. and Wang, H.** (1993). Mixed-discrete nonlinear optimisation with simulated annealing. Engineering Optimisation, 21(4): 277–291, 1993. ISSN 0305-215X.
- [95] **Cao, Y. J. and Wu, Q. H.** Mechanical design optimisation by mixed-variable evolutionary programming. Proceedings of the 1997 IEEE Conference on Evolutionary Computation, pp. 443–446.
- [96] **Lin, S., Zhang, C. and Wang, H.** On mixed-discrete nonlinear optimisation problems: A comparative study. Engineering Optimisation, 23(4): 287–300, 1995. ISSN 0305-215X.
- [97] **Wu, S.-J. and Chow, P.-T.** Genetic algorithms for nonlinear mixed discrete-integer optimisation problems via meta-genetic parameter optimisation. Engineering Optimisation, 24(2): 137–159, 1995. ISSN 0305-215X.
- [98] **Lampinen, J. and Zelinka, I.** Mixed integer-discrete-continuous optimisation by Differential Evolution, Part 1: the optimisation method. In: Ošmera, Pavel (ed.) (1999). Proceedings of MENDEL'99, 5th International Mendel Conference on Soft Computing, June 9.–12. 1999, Brno, Czech Republic. Brno University of Technology, Faculty of Mechanical Engineering, Institute of Automation and

- Computer Science, Brno (Czech Republic), pp. 71–76. ISBN 80-214-1131-7.
- [99] **Siddal, J.N.** Optimal engineering design: principles and applications. Mechanical engineering series/ 14. Marcel Dekker Inc. 1982. ISBN 0-8247-1633-7.
 - [100] **Chen, J. L. and Tsao, Y. C.** Optimal design of machine elements using genetic algorithms. Journal of the Chinese Society of Mechanical Engineers, 1993, 14(2), pp. 193-199.
 - [101] **Duym, S., Stiens, R. and Reybrouck, K.** Evaluation of Shock Absorber Models, Vehicle System Dynamics, Vol. 27, 1997, pp. 109-127.
 - [102] **Mollica, R. and Youcef-Toumi, K.** A nonlinear Dynamic Model of a Monotube Shock Absorber, Proceeding of the American Control Conference Albuquerque, New Mexico June 1997.
 - [103] **Lang, H.H.** A Study of the Characteristics of Automotive Hydraulic Dampers at High Stroking Frequencies, 1977, Ph.D Thesis, Department of Mechanical Engineering, University of Michigan.
 - [104] **Segel, L. and Lang, H.H.** The Mechanics of automotive hydraulic dampers at High Stroking Frequency, 1981, Proceedings of the 7th IAVSD Symposium on the Dynamics of Vehicles, Cambridge.
 - [105] **Reybrouck, K.** A nonlinear Parametric Model of an Automotive Shock Absorbers, Preprint, Monroe Europe, Industriezone 1, 3800 Sint-Truiden, Belgium.
 - [106] **Wallaschek, J.** Dynamics of Nonlinear Automotive Shock Absorbers, 1990, International Journal of Nonlinear Mechanics, Vol.23, pp.299-308.
 - [107] **Hall, B. B. and Gill, K. F.** Performance of a telescopic dual-tube automotive damper and the implications for vehicle ride prediction, Proceeding of ImechE, part D, 200 (D2), pp. 115-123, 1986.
 - [108] The MathWorks, Inc, Using MATLAB.
 - [109] **Worden, K. and Tomlinson, G. R.** Parametric and Non-parametric identification of automotive shock absorber, Proceeding of the 10th International Modal Analysis Conference, San Diego, pp. 764-771, 1992.
 - [110] **Surace, C., Worden, K. and Tomlinson, G. R.** On the nonlinear characteristics of automotive shock absorbers, Proceedings, of IMechE, part D, 206, pp. 3-16, 1992.
 - [111] **Cafferty, S.** Characterization of automotive shock absorbers using time and frequency domain techniques, Thesis, School of Engineering, Victoria University

of Manchester, 1996.

- [112] **Cafferty, S., Worden, K. and Tomlinson, G. R.** Characterization of automotive shock absorbers using random excitation, Proceedings, of IMechE, part D, 209, pp. 239-248, 1995.
- [113] **Surace, C., Storer, D. and Tomlinson, G. R.** Characterizing an automotive shock absorber and the dependency on the temperature, Proceedings of the 10th International Model Analysis conference, San Diego, pp. 1317-1326, 1992.
- [114] **Karadayi, R. and Masada, G. Y.** A nonlinear shock absorber model, Proceeding of the Symposium on Simulation and Control on Ground Vehicle and Transportation Systems, pp. 149-165, 1986.
- [115] **Besinger, F. H., Cebon, D. and Cole, D. J.** Damper models for heavy vehicle ride dynamics, Vehicle System Dynamics, Vol. 24, 1995, pp. 35-64.
- [116] **Allen, D. M.,** Mean Square Error of prediction as a criterion for selecting variables, Technometrics, Vol. 13, No. 3, 1971, pp. 469-475.
- [117] **R. S., Sharp and S. A., Hassan** The fundamental of passive automotive suspension system design.
- [118] **R. S., Sharp and D. A., Crolla** Road Vehicle Suspension System Design – a review, Vehicle System Dynamics, Vol. 16, pp. 167-192.
- [119] **T., Dahlberg** Optimization Criteria for Vehicles Travelling on a randomly Profile Road – a survey, Vehicle Systems Dynamics, Vol. 8, No. 4, September 1979, pp. 239-252.
- [120] **Gen, M. and Cheng, R.** Genetic Algorithms and Engineering Optimization. John Wiley & Son, Inc. 2000.
- [121] **Ishibasi, H., Aguirre, H. E., Tanaka, K. and Sugimura, T.** Multi-objective Optimization with Improved Genetic Algorithm. IEEE 2000, pp. 3852-3857
- [122] **Mingqiang, L., Jisong, K. and Lin, D.** GA-based Multi-objective Optimization. Proceeding of the 3rd World Congress on Intelligent Control and Automation. IEEE 2000, pp. 637-640.
- [123] **Kalman, R. E. and Bucy, R.** New results in Linear Filtering and Prediction, Journal of Basic Engineering (ASME), Vol. 83D, 1961, pp. 95-108.
- [124] **Gelb A.** Applied Optimal Estimation, 1974, The Analytic Sciences Corporation.
- [124] **Lewis F. L.** Optimal Estimation with Introduction to Stochastic Control Theory, 1986, John Wiley&Sons, Inc.

- [126] Tsi Propac, Techno-Science, Inc. 1994.
- [127] **Pedchote, C. and Purdy, D. J.**, Parameter Estimation of a Quarter-Vehicle Suspension Using Hybrid Differential Evolution, Proceedings of the Institution of Mechanical Engineering, Part D.

APPENDIX A Parameter Estimation of an Adjustable Automotive Damper.

A1. Introduction

In the investigation carried out in Chapter 3, the damper was modelled using ideal element such as a dashpot or combinations of dashpot and spring elements. The analysis of results has shown that Model 3, where a bilinear dashpot connected in series with a linear compliance spring, was the ‘best’ model among those models considered. The model gave a quality of fit in term of MSE value of less than 1%, and provided the best compromise between time consumed in the estimation process and the sufficiency of describing the dynamic behaviour of the test damper in the frequency range of interest, 0.5-30 Hz.

Though Model 3 was able to capture dominant dynamic behaviour of the test damper, which was the bilinear damping characteristics and compliance properties, it however did not include other important non-linear effects such as friction and hysteresis. Parameter estimation of a more complex damper model, which included such non-linear effects and was thus capable of describing damper behaviour better, was carried out and the results are presented in this section. Though the parameter estimation was applied to a different automotive damper, the objective of the investigation remained the same. The reasons were not only to appreciate the different damper design configuration but also to compare the estimation results obtained by other researchers. The prime objective was to examine the performance of the dvHDE against different numerical search methods when applied to parameter estimation of a more complex, realistic damper model.

One particular work by Purdy [1], where the dynamics of an adjustable damper were examined both experimentally and theoretically, has been selected for the comparison of the estimation result. In [1] a non-linear parametric model of a commercially available, gas-pressurised adjustable automotive damper was developed based on its detailed physical construction. The model included frictional effect, the compressibility of the fluid, trapped gas and expansion of the cylinder. A diagrammatic representation of the adjustable damper is provided in Fig. A1. For more details on the working operations of the damper and the development of mathematical equations describing the model behaviour, the readers are referred to [1] and the literature therein. The interest here is limited to the aspect of the model parameter estimation.

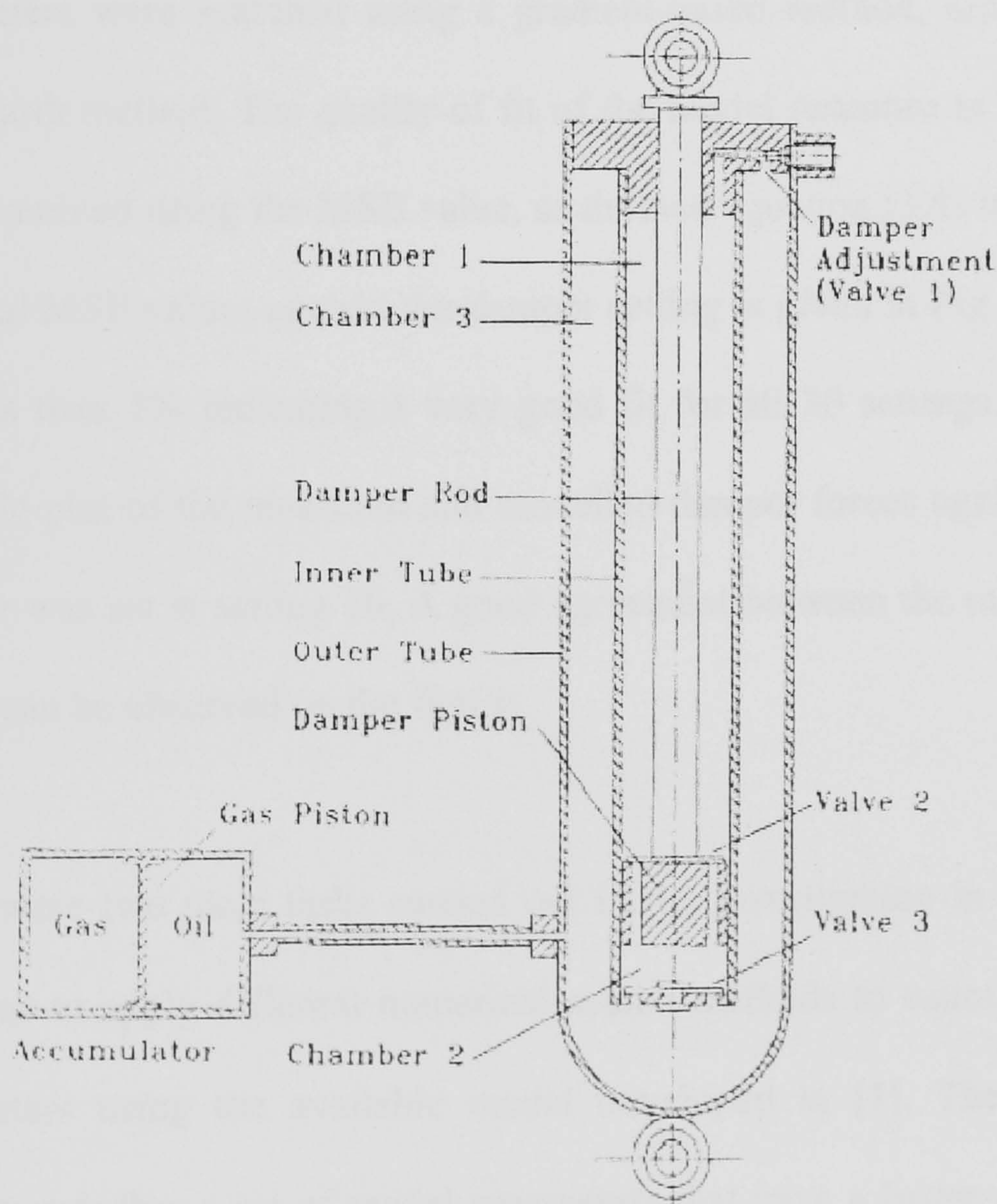


Fig. A1 Diagrammatic representation of the adjustable damper, taken from [1].

The adjustable damper had 20 discrete settings that modified both bump and rebound characteristics. Its model had been formulated with five tuneable parameters; mass flow coefficients for valve 1, valve 2, and valve 3, respectively denoted by R_{m1} , R_{m2} and R_{m3} , effective bulk modulus denoted by K_e and friction denoted by f_f . In the estimation of the model parameters the mass flow coefficient for valve 3, R_{m3} , and friction, f_f , were fixed, thus there were only three parameters to be identified. The values of R_{m3} was fixed at 2.597 MN s/m² kg, approximately 15 percent below the value at which cavitation occurred. The value of the kinetic friction did not vary much with the damper settings, the value of f_f was thus fixed at 50 N. The estimation technique used was a Least Squares technique, which minimised quadratic norm of the difference between the measured and modelled damper forces. The ‘best’ model parameters were searched using a gradient-based method, namely the Levenberge-Marquardt method. The quality of fit of the model response to the experimental data was examined using the MSE value, as defined equation (3.9) in Chapter 3. A plot of the final MSE values against the damper setting is given in Fig. A2. The MSE values are less than 1% indicating a very good fit for all 20 settings. Fig. A3 provides an example plot of the measured and modelled damper forces against velocity when the damper was set at setting 10. A good agreement between the measured and modelled forces can be observed on the figure.

There were two main tasks carried out in the investigation in this section. The first task was to apply different numerical search methods to estimate the damper model parameters using the available model developed in [1]. The second task was to examine whether a set of model parameters that gave a better quality of fit could be found and/or the time consumed in the estimation process could be reduced. The

numerical methods considered for the comparison were the gradient-based Levenberge-Marquardt method, the DS, DE and dvHDE methods. The parameter estimation procedure was the same as in [1], which was to minimise the quadratic norm of the difference between the measured and modelled forces in the Least Square sense, and with three parameters to be identified; R_{m1} , R_{m2} and K_e .

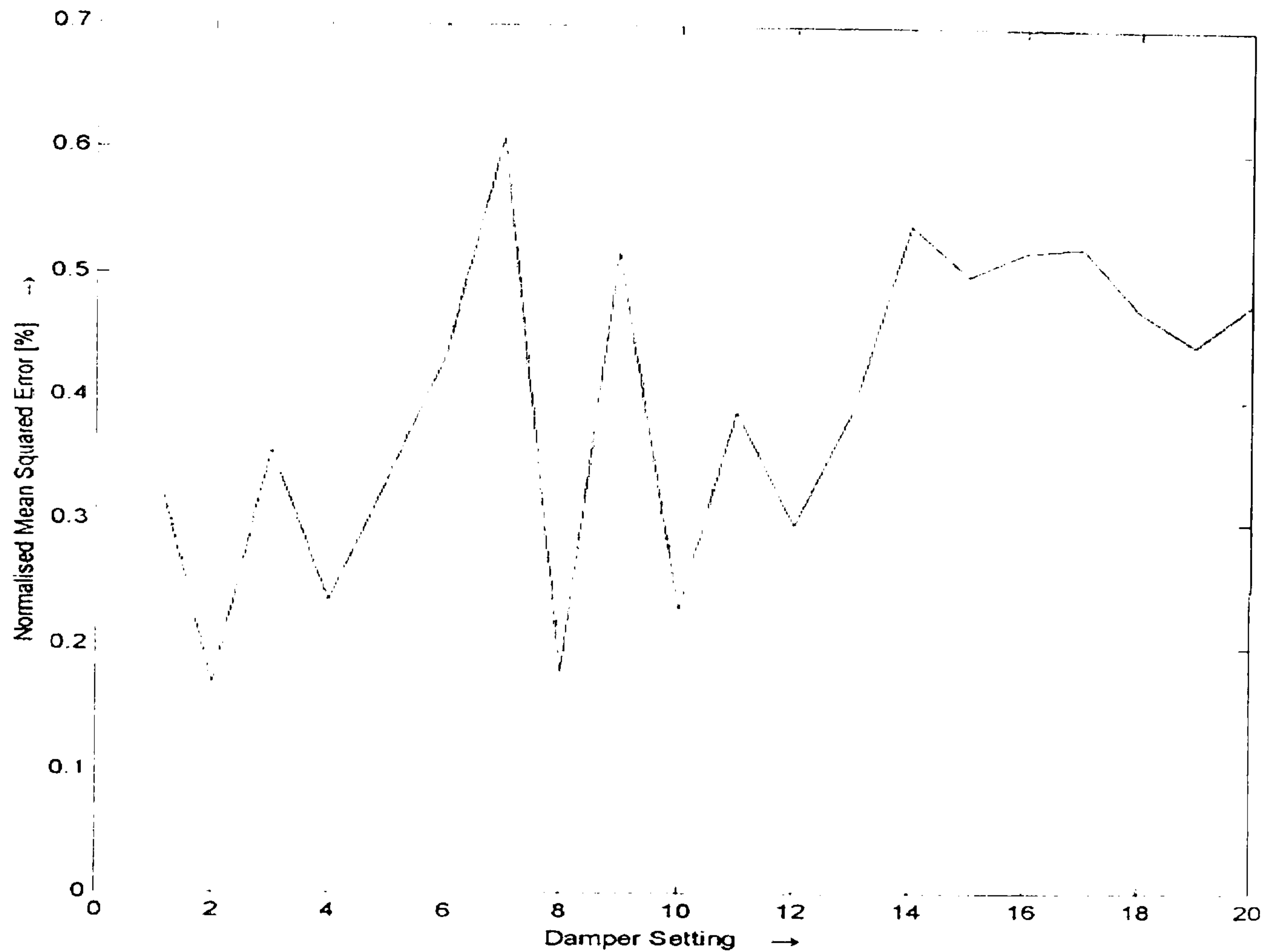


Fig. A2 Plot of the MSE values against the damper setting, taken from [1].

The lower and upper parameter bounds for the four numerical search methods were initially set as follows; the lower bound = $[10 \times 10^3, 1 \times 10^6, 100 \times 10^6]$, the upper bound = $[40 \times 10^3, 70 \times 10^6, 600 \times 10^6]$. These values were taken from the estimated parameters reported in [1]. The initial starting points for the gradient-based and the DS methods were randomly selected between the lower and upper bounds. The two methods were set-up to terminate when maximum iteration of 1000 or maximum function evaluations of 2000 was reached. The control parameters of the DE and dvHDE

methods were set at the same following values; population size $n_i = 15$, mutation factor $\mu_m = 0.7$, crossover factor $\mu_c = 0.8$, and terminated when maximum iteration reached 150 generations.

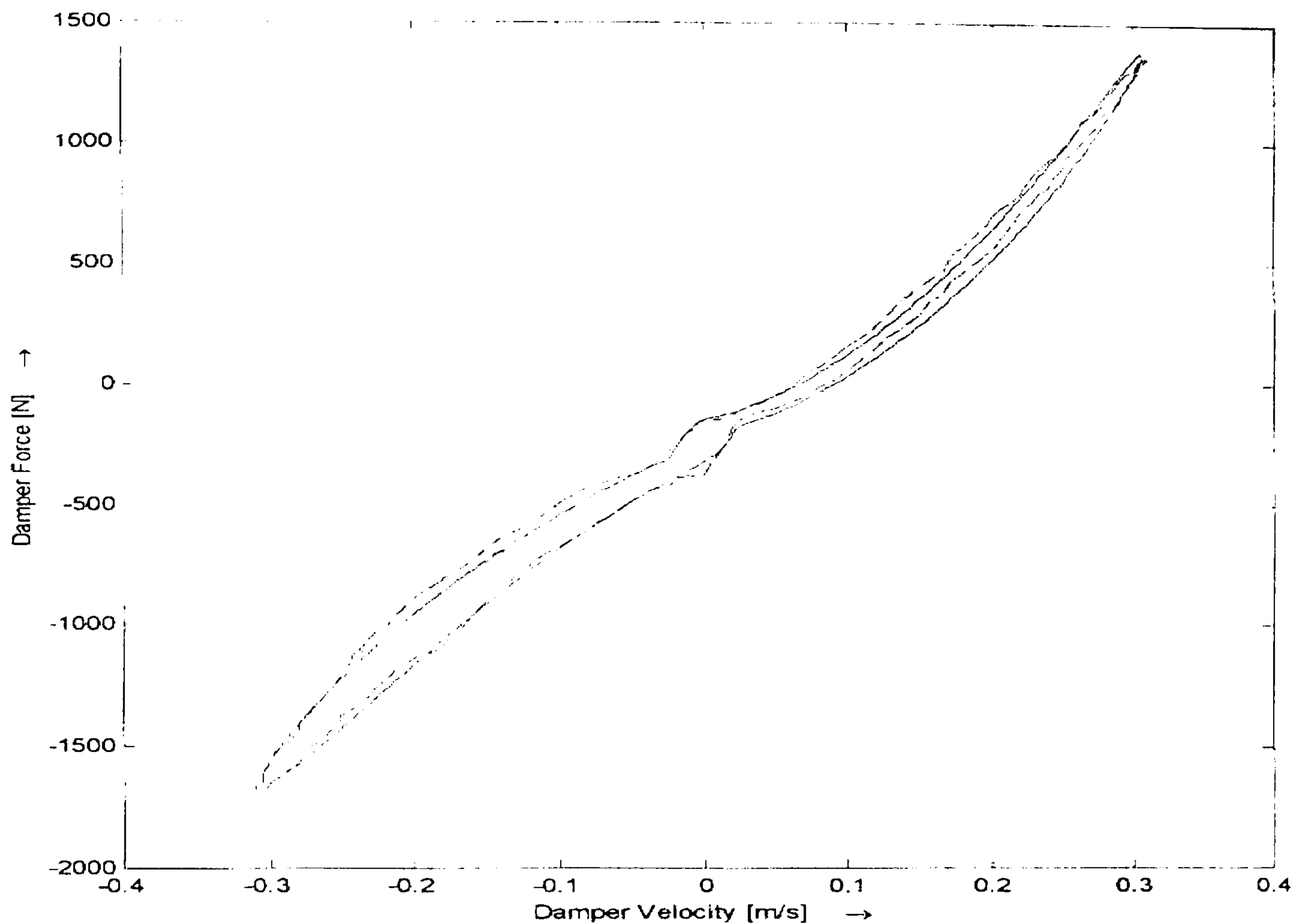


Fig. A3 Plot of the damper force against velocity for setting 10;
— theoretical, ---experimental.

The performance of each numerical search method was compared by examining the MSE values, number of function evaluations and the CPU-time consumed. The results are presented in Table A1 and Table A2. For each setting, the values in the tables are the mean and standard deviation from 20 independent runs of the MSE values, denoted by $m_{J_{MSE}}$ and $\alpha_{J_{MSE}}$, and of the CPU-time, denoted by $m_{T_{CPU}}$ and $\alpha_{T_{CPU}}$ respectively. The number of function evaluations was useful in the analysis and comparison of the results. It was however not included in the tables, as its value was proportional to and could be reflected by the value of CPU-time.

A2. Analysis of results

First considering the results obtained from the gradient-based Levenberge-Marquardt method (referred here as the GB method), the results are presented in Table A1. The time consumed for most of the GB's runs were small, and were occasionally large. These were reflected by small values of $m_{T_{CPU}}$ and high values of $\alpha_{T_{CPU}}$. Though the consumed CPU-time were small, there were only 28 out of 400 runs that the MSE values were less than 5%. This means the GB method had less than 7 percent success in finding the model parameters that gave a good quality of fit. The final MSE values depended on the initial starting points and were often large. These resulted in high values of $m_{J_{MSE}}$ and $\alpha_{J_{MSE}}$ in Table A1. The results suggested the GB method suffered severely from local minima problem. The results similar to that in Fig. A2 could only be obtained when careful trial-and-error procedure, which re-assigned the lower and upper parameter bounds and provided the 'right' starting points for each setting, were exercised. To provide the search algorithm with good parameter interval and a 'right' initial starting point was not either easy or straightforward, and, in fact, was time consuming.

The performance of the DS method also depended on the initial starting points. Like the GB method, the DS encountered problem of obtaining local minima. The problem was however less severe. For each setting, there were both high and low MSE values. In total there were 242 runs that the MSE values were less than 5%. This means the DS method had a success rate of finding the optimum points of less than 60 percent. When not encountering local minima problem, the method had a very slow convergence rate, and terminated when maximum number of function evaluations set

was reached. When this occurred, the CPU-time required was large, see Table A1. The mean and standard deviation of both the MSE values and CPU-time for all settings were high. This was because the algorithm experienced a mixture of the two problems; obtaining local minima and having slow convergence rate problems.

	GB method				DS method			
	Mean values		Standard deviations		Mean values		Standard deviations	
	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$\alpha_{J_{MSE}}$	$\alpha_{T_{CPU}}$ [s]	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$\alpha_{J_{MSE}}$	$\alpha_{T_{CPU}}$ [s]
Setting 1	123.4	40.9	97.6	86.9	65.3	803.4	84.4	689.7
Setting 2	99.0	72.4	62.1	121.9	31.6	879.7	43.7	725.9
Setting 3	128.3	99.3	75.7	136.5	41.3	1053.6	48.0	904.3
Setting 4	105.2	37.7	78.3	77.3	47.9	395.5	90.5	470.6
Setting 5	55.8	38.6	41.2	82.6	18.8	826.5	26.2	713.5
Setting 6	106.1	98.7	59.6	133.0	17.3	611.2	21.0	586.5
Setting 7	82.0	40.7	54.8	86.9	22.5	992.4	33.5	779.1
Setting 8	35.2	74.9	26.3	126.2	26.7	756.5	12.1	647.2
Setting 9	52.2	94.9	28.0	130.2	26.3	775.5	77.6	698.2
Setting 10	29.2	45.8	18.3	101.8	31.9	830.1	85.6	686.8
Setting 11	49.7	11.5	40.3	1.1	31.5	723.0	64.8	705.9
Setting 12	60.7	12.5	30.7	3.6	48.2	470.0	80.0	543.9
Setting 13	29.2	130.2	25.5	146.4	30.8	367.9	41.5	442.8
Setting 14	58.7	108.8	49.8	149.0	39.7	409.9	53.3	303.4
Setting 15	58.3	11.9	44.9	1.6	43.0	417.4	83.6	417.1
Setting 16	45.4	40.5	33.0	88.6	47.4	407.4	118.1	338.3
Setting 17	25.7	69.1	14.4	115.2	36.3	278.7	105.8	272.9
Setting 18	13.1	11.7	8.5	2.1	12.7	605.2	22.3	404.6
Setting 19	33.9	45.6	24.8	100.9	11.5	570.6	26.8	427.8
Setting 20	26.9	129.5	22.7	184.5	20.2	491.6	34.4	481.7

Table A1 Estimation results by the GB and DS methods.

	DE method				dvHDE method			
	Mean values		Standard deviations		Mean values		Standard deviations	
	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$\alpha_{J_{MSE}}$	$\alpha_{T_{CPU}}$ [s]	$m_{J_{MSE}}$	$m_{T_{CPU}}$ [s]	$\alpha_{J_{MSE}}$	$\alpha_{T_{CPU}}$ [s]
Setting 1	0.158	700.9	0.0037	303.4	0.158	337.1	0.0032	80.7
Setting 2	0.150	527.7	0.0019	316.3	0.150	261.6	0.0015	57.2
Setting 3	0.151	668.2	0.0057	409.1	0.151	308.5	0.0056	111.1
Setting 4	0.154	828.5	0.0097	468.5	0.151	327.4	0.0074	146.3
Setting 5	0.150	566.4	0.0044	247.5	0.151	321.8	0.0063	127.3
Setting 6	0.154	525.2	0.0078	328.8	0.152	268.2	0.0057	137.9
Setting 7	0.154	384.4	0.0029	49.4	0.155	290.7	0.0053	90.9
Setting 8	0.163	494.8	0.0061	245.9	0.162	248.8	0.0051	47.7
Setting 9	0.167	448.6	0.0042	139.9	0.169	251.3	0.0067	66.6
Setting 10	0.184	368.2	0.0002	33.5	0.186	196.6	0.0039	56.0
Setting 11	0.214	449.2	0.0135	248.2	0.210	232.6	0.0002	34.6
Setting 12	0.257	395.8	0.0239	217.7	0.249	213.7	0.0001	44.7
Setting 13	0.292	331.1	0.0001	53.7	0.292	216.6	0.0001	24.8
Setting 14	0.381	356.8	0.0572	180.9	0.362	194.2	0.0001	26.8
Setting 15	0.415	267.3	0.0001	17.0	0.415	169.5	0.0001	20.6
Setting 16	0.486	247.1	0.0000	20.1	0.486	164.7	0.0000	31.4
Setting 17	0.508	253.3	0.0000	25.6	0.508	151.5	0.0000	13.4
Setting 18	0.474	236.4	0.0000	6.8	0.474	160.5	0.0000	26.3
Setting 19	0.445	227.1	0.0000	9.8	0.445	147.2	0.0000	9.6
Setting 20	0.480	220.9	0.0000	10.9	0.480	151.0	0.0000	20.6

Table A2 Estimation results by the DE and dvHDE methods.

The DE and dvHDE methods were able to find the model parameters that gave the MSE value of less than 1% for all 400 runs. The final MSE values for all settings were less than, if not equal to, that provided in Fig. A1. This means, firstly, the two methods consistently found the optimum points without suffering premature

convergence problem. Secondly, the methods have found the parameter sets that gave better qualities of fit for the model being studied. With the initial parameter lower and upper bounds set, the dvHDE method required the CPU-time of approximately 100 seconds to find the optimum points, while the DE method required the time in an interval of 200-700 seconds. A close analysis revealed that the estimated parameters for several settings were very close to the upper limit. This suggested that the parameter upper bound was reached. The investigation was further carried out with a much wider search space. The parameter lower bound remained the same as before, and the upper bound was increased to $[5 \times 10^5, 5 \times 10^8, 5 \times 10^9]$. With the new parameter bounds, the performance of the GB and DS methods became worse because of a larger number of local minima present and slower convergence rate due to a larger search space. The DE and dvHDE, again, did not encounter local minima problem. Instead, the methods found the parameter sets that gave lower MSE values than that when the previous parameter bounds were used. Further reduced MSE values were obtained in the expense of more CPU-time required as expected. The time was in between approximately 150-300 seconds for the dvHDE method, and 200-800 seconds for the DE method, see Table A2. Comparing the values of $m_{J_{MSE}}$ and $\alpha_{J_{MSE}}$ of the two methods, they are very close to each other. It can be said that the two methods gave the same results. The important differences are in the values of $m_{T_{CPU}}$ and $\alpha_{T_{CPU}}$. The mean values $m_{T_{CPU}}$ of the dvHDE method are smaller than that of the DE method for all 20 settings. Comparing how the required CPU-time of two methods varied, the $\alpha_{T_{CPU}}$ values by the dvHDE were small, and smaller than that of the DE for most of the settings. From the values of $m_{T_{CPU}}$ and $\alpha_{T_{CPU}}$, it can be said that the dvHDE method consistently required less time than the DE method. The mean values

$m_{J_{MSE}}$ and $m_{T_{CPU}}$ in Table A2 are plotted in Fig. A4. From parts (a) and (b) of the figure, the discussions made earlier became clear; that was the MSE values were very close to each other, and the dvHDE consumed less amount of time. The figure also shows the number of function evaluations in Fig. A4(c). It can be seen that the required number of function evaluations were approximately 900 by the dvHDE method and approximately 1700 by the DE method. Less number of function evaluations were reflected in less time consumed. Therefore, based on the CPU-time required to obtain the optimum points, the dvHDE method performed better over the DE method by 20-60%.

A3. Summary

The model parameters of an adjustable damper were estimated using four different numerical search methods. The investigation has shown that, first, the gradient-based method can suffer greatly with the problem of obtaining local minima. Acceptable estimation results can only be obtained after careful and time consuming trial-and-error process is exercised in order to provide the search algorithm with ‘proper’ lower and upper bounds and ‘good’ initial starting points. Secondly, the DS method encountered two main problems. They were mis-convergence due to local minima and slow convergence rate problems. Thirdly, the use of the population-based such as the DE and dvHDE methods relaxed the requirement on the priori knowledge of the parameter lower and upper bounds, and the initial starting point for the search process.

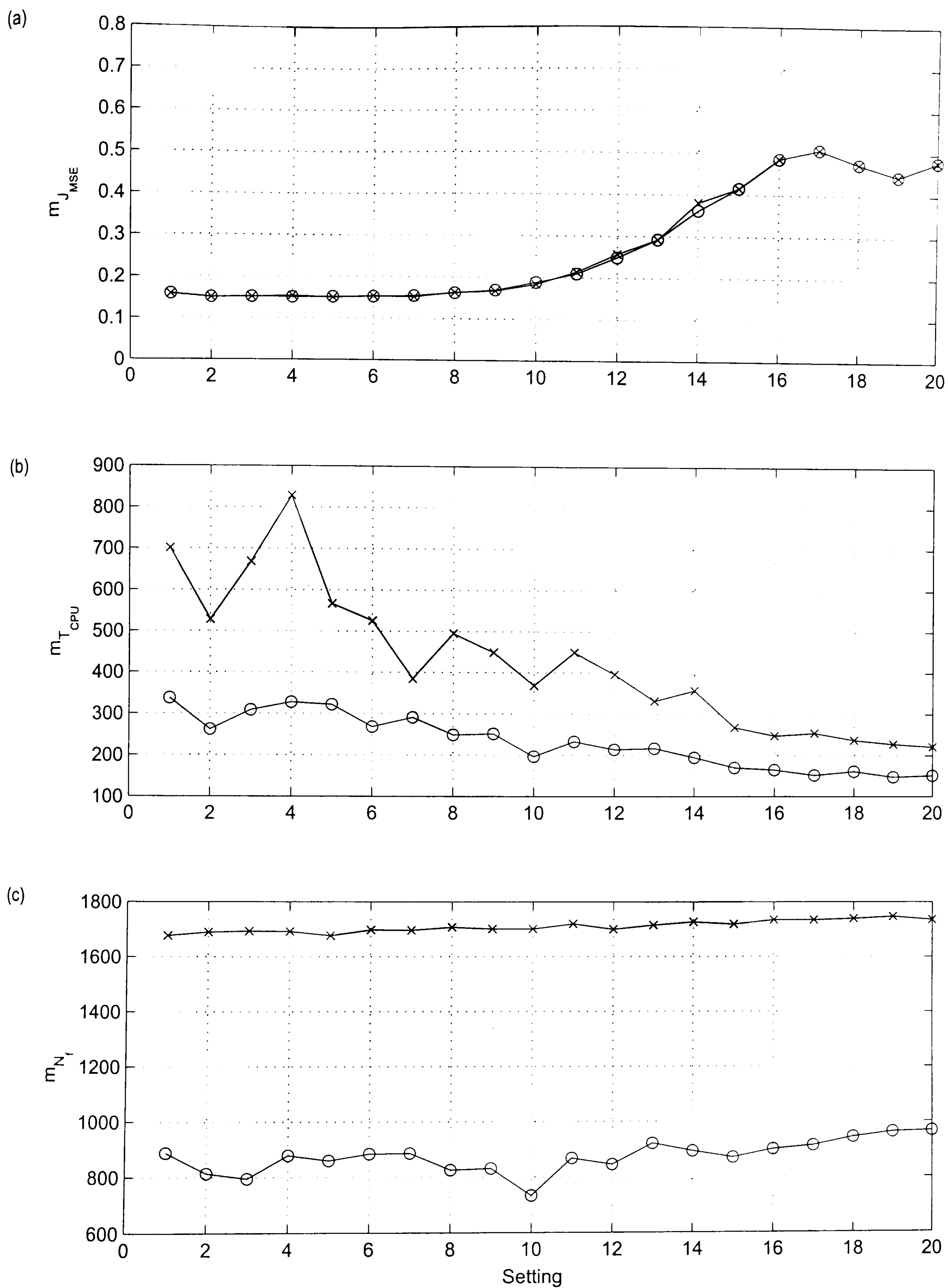


Fig. A4 Plot showing the performance of the DE and dvHDE method:
 -x-x- DE method, -o-o- dvHDE method.

The methods not only consistently found the optimum points, but they also found different sets of model parameters that gave a better quality of fit. This could possibly lead to different discussion and conclusions of the investigation into the model behaviour. Finally, it has been shown that the DE and dvHDE methods gave the same final quality of fit. The dvHDE however required less number of function evaluations and less amount of time. In term of the CPU-time required, the dvHDE method performed better over the DE method by up to 60%. The method was therefore the best among the four different numerical search methods considered. The parameters of a more realistic, complex damper model have therefore been successfully identified. The use of the dvHDE method offers great advantages of able to find the optimum point with high percentage of success, and achieving it in a short period of time.

APPENDIX B Vehicle Axis Systems

The studies of vehicle dynamics are concerned with the vehicle movements such as accelerating and braking, ride and turning. The equations of motion employed in the studies are related to a set of body-fixed axes. This axis system is right handed. The positive sense of rotation is clockwise when viewed from the origin along the positive direction of the axis. Fig. B1 shows the system. In simple studies of whole vehicle braking, accelerating and turning analyses, it is appropriate to position the origin of the axes at the center of mass of the total vehicle in the case of cars [3]. However, in ride analysis and in suspension design when the movement of vehicle body (the sprung mass) relative to the other moving parts of the suspension and wheels (the unsprung masses), it is usual to place the vehicle axis stem at the CG of the sprung mass.

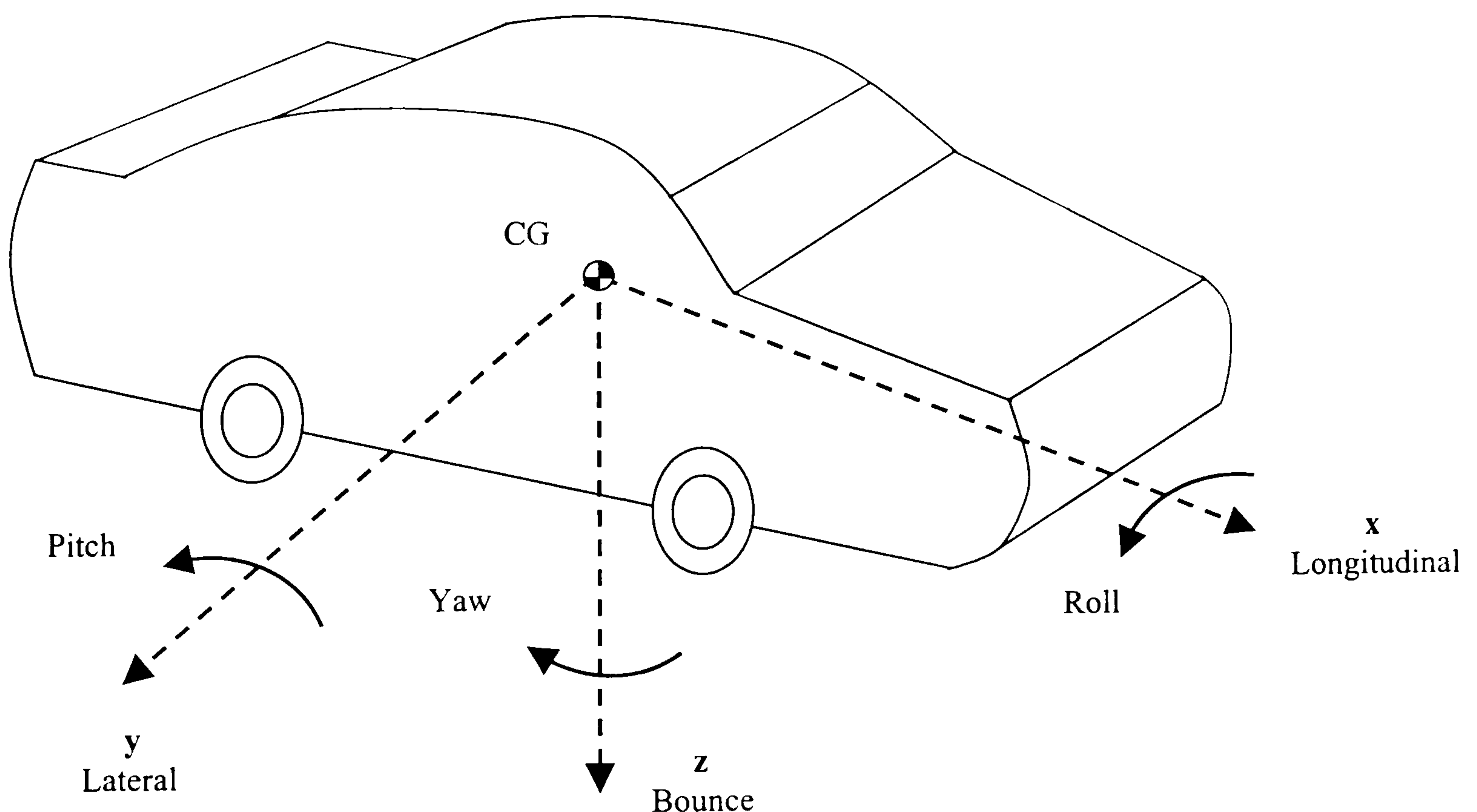


Fig. B1 SAE vehicle axis system.

Fig. B1 shows the vehicle fixed coordinate system according Society of Automotive Engineering (SAE) convention. See [3] for a comprehensive list of SAE vehicle dynamics terminology. Three vehicle motion modes, which are of the prime concerns in this research work, are roll, pitch and bounce. The positive directions are taken as shown in Fig. B1.

APPENDIX C

This appendix presents some preliminary results when the Landrover under test was modelled using a four degrees of freedom (4DOF) model that takes into account the effect of the vehicle suspension linkages, referred to as Model 2, and a seven degrees of freedom (7DOF) model, referred to as Model 3. First the models are described and then the quality of the models in capturing the dynamics of the vehicle are discussed and compared to Model 1 of the main thesis Chapter 5. It must be born in mind that the estimation of the model parameters was carried out in the frequency domain using the dvHDE search method, and due to a limited amount of time, the discussion of the estimated results are based on a limited number of runs. The estimation in the time domain and a complete investigation into the different numerical search methods when applied to the parameter estimation of Model 2 and Model 3 are left for future work.

C1. Model 2: vehicle bounce and pitching model with linkages

In Model 2, the vehicle is modelled as three lump masses; the sprung mass, m_s , the front and rear unsprung masses, m_f and m_r , respectively. The sprung mass, m_s , which represents the vehicle body, is allowed to bounce and pitch about its centre of gravity, and having the moment of inertia of I_s . The front and rear unsprung masses, m_f and m_r , are connected to the sprung mass by two revolute joints, as shown in Fig. C1. The unsprung masses are allowed to rotate about their joints relative to the sprung mass, thus each unsprung mass contributing one degree of freedom. The centres of mass of

the front and rear unsprung masses are located at the ends of the two linkages, at which the weights of the axle and wheel hardware are concentrate. The model is therefore a 4DOF, planar model in the longitudinal and vertical planes.

At the front axle, the compliance of the suspension springs, dampers and bushes are modelled by a linear spring connected in parallel with a bilinear viscous damping element. The two elements are connected vertically between the sprung and unsprung mass.

At the rear axle, the suspension is modelled as a linear spring connected vertically relative to the sprung mass, and in parallel with an inclined bilinear damping element. The front and rear tyres are initially modelled as linear springs.

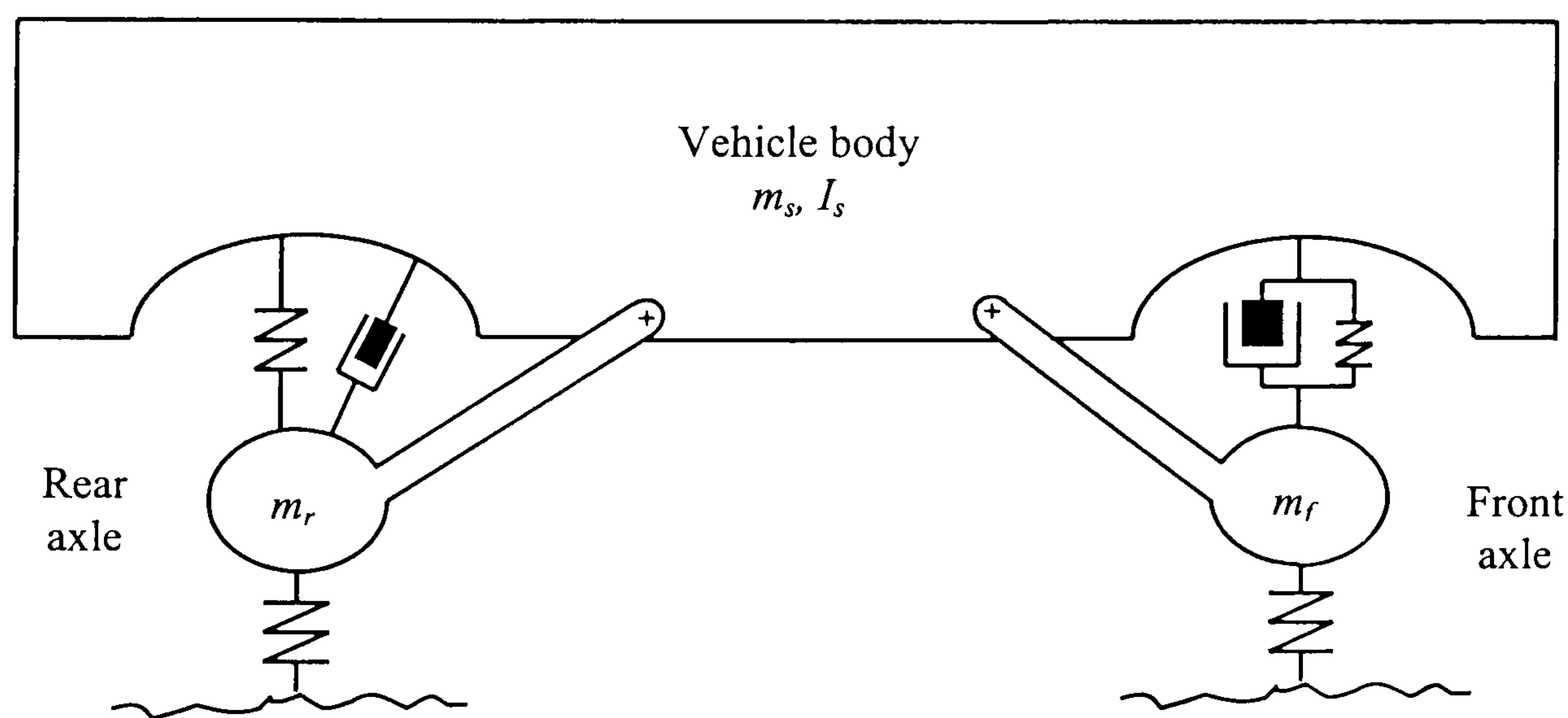


Fig. C1 A four degrees of freedom (4DOF) model with linkages (Model2)

Model 2 is nonlinear and contains 12 parameters to be identified. With the two linkages, derivation of the model equations of motion by hand becomes too tedious. As a consequence, the generation of the equations of motion for Model 2 requires

resorting to computer software. In this work, a commercially available software tool named TSi Propac [125] is used. The software provides the model equations of motion in a C-language code file, which is in turn compiled and used in Matlab and Simulink to obtain the model responses. With the software, a more realistic model that allows not only the linkages but also the inclined dampers to be included relatively conveniently. For more details on how a model can be constructed using TSi Propac, the reader is referred to the reference [125].

C2. Model 3: a 7DOF vehicle model

Model 3 consists of three lump masses; the sprung mass representing the vehicle body, the front and rear unsprung masses representing the front and rear axles and wheel hardware, respectively. The model is depicted in Fig. C2. Model 3, which is an extended version of Model 1, takes into account the vehicle body bounce, pitch and roll, as well as the front and rear axles bounce and roll. The model is thus a 7DOF model.

The front suspension, both the far-side and near-side, are modelled as a linear spring connected in parallel with a bilinear viscous damping element, and the tyre is simply modelled as a linear spring. Similar to Model 1, by applying Newton's second law, the equations of motion of the sprung and unsprung masses can be obtained. The derivation of this is provided below. The model responses are then obtained by simulation using Matlab and Simulink.

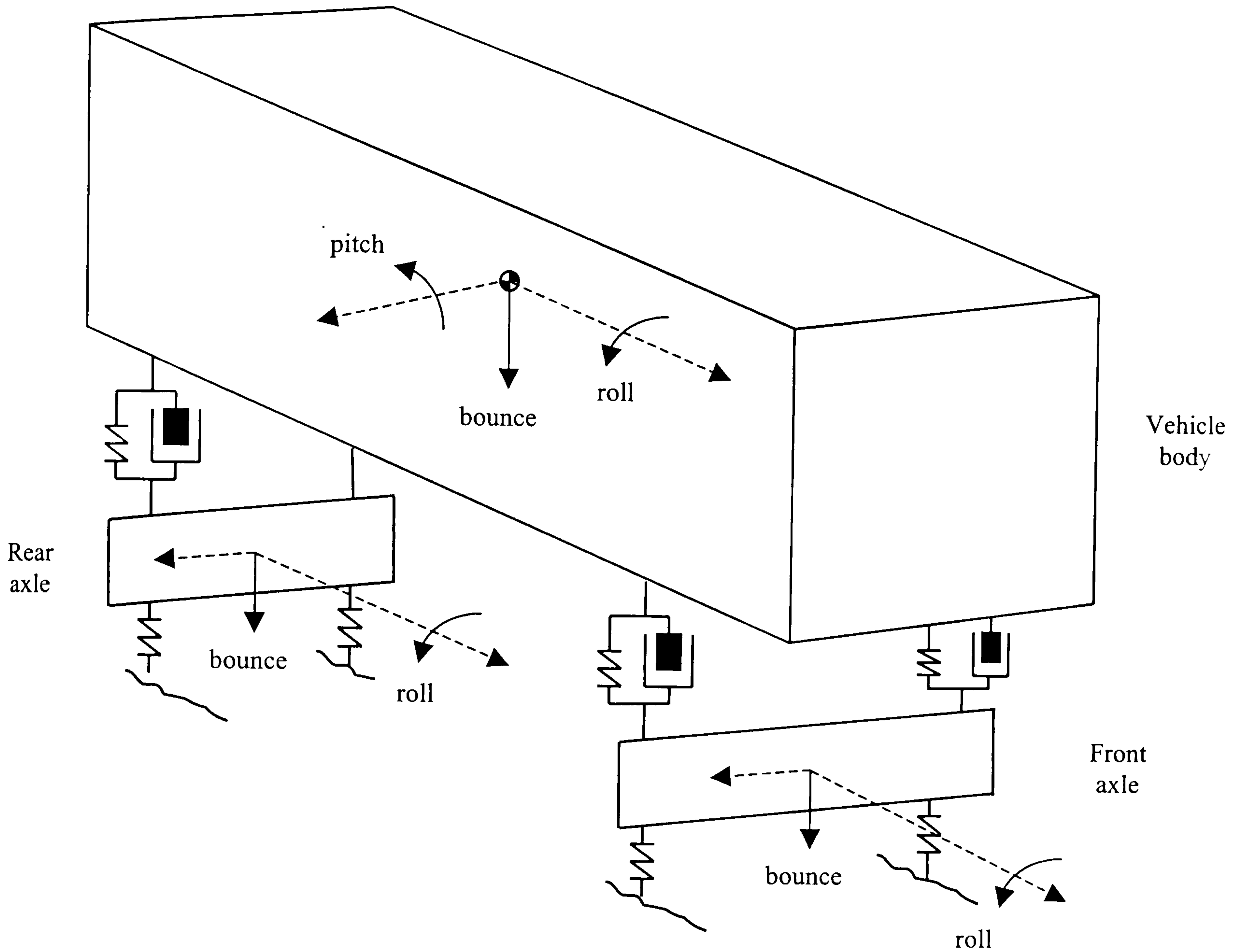


Fig. C2 A seven degrees of freedom (7DOF) vehicle model

The derivation of the model equations of motion are provided below and the following notations are used.

d_{f1}, d_{f2}	front far-side, and near-side damping coefficients
d_{r3}, d_{r4}	rear near-side, and far-side damping coefficients
f_{d3}, f_{d4}	rear near-side and far-side suspension damping forces
f_{s1}, f_{s2}	front far-side and near-side suspension forces
f_{s3}, f_{s4}	rear near-side and far-side suspension spring forces

f_{yf1}, f_{yf2}	front far-side and near-side tyre forces
f_{yr3}, f_{yr4}	rear near-side and far-side tyre forces
g	gravitational constant
I_{fxx}, I_{rxx}	moment of inertia of the front and rear unsprung masses about longitudinal axis
I_{sxx}, I_{syy}	moment of inertia of the sprung mass about longitudinal and lateral axes
k_{sf1}, k_{sf2}	front far-side, and near-side suspension spring stiffnesses
k_{sr3}, k_{sr4}	rear near-side, and far-side suspension spring stiffnesses
k_{yf1}, k_{yf2}	front far-side, and near-side tyre stiffnesses
k_{yr3}, k_{yr4}	rear near-side, and far-side tyre stiffnesses
m_s	sprung mass
m_f, m_r	front and rear sprung mass
x_n	longitudinal distance from the centre of gravity of location ‘n’ where force is applied
y_n	lateral distance from the centre of gravity of location ‘n’ where force is applied
z_{gf1}, z_{gf2}	front far-side and near-side ground input displacements
z_{gr3}, z_{gr4}	rear near-side and far-side ground input displacements
$z_f, \dot{z}_f, \ddot{z}_f$	displacement, velocity, and acceleration of the front unsprung mass
$z_r, \dot{z}_r, \ddot{z}_r$	displacement, velocity, and acceleration of the rear unsprung mass
$z_s, \dot{z}_s, \ddot{z}_s$	displacement, velocity, and acceleration of the sprung mass
$\theta_{xf}, \dot{\theta}_{xf}, \ddot{\theta}_{xf}$	roll displacement, velocity, and acceleration of the front unsprung mass

$\theta_{xr}, \dot{\theta}_{xr}, \ddot{\theta}_{xr}$ roll displacement, velocity, and acceleration of the rear unsprung mass

$\theta_{xs}, \dot{\theta}_{xs}$ roll displacement and velocity of the sprung mass

$\theta_{ys}, \dot{\theta}_{ys}$ pitch displacement and velocity of the sprung mass

Noting that:

1. The numbers at different places on the masses indicate the locations where forces are applied.
2. x_n and y_n , the longitudinal and lateral position of location 'n', are taken to be positive in the positive direction respect to the origin.

Considering a free-body diagram for the sprung mass in Fig. C3;

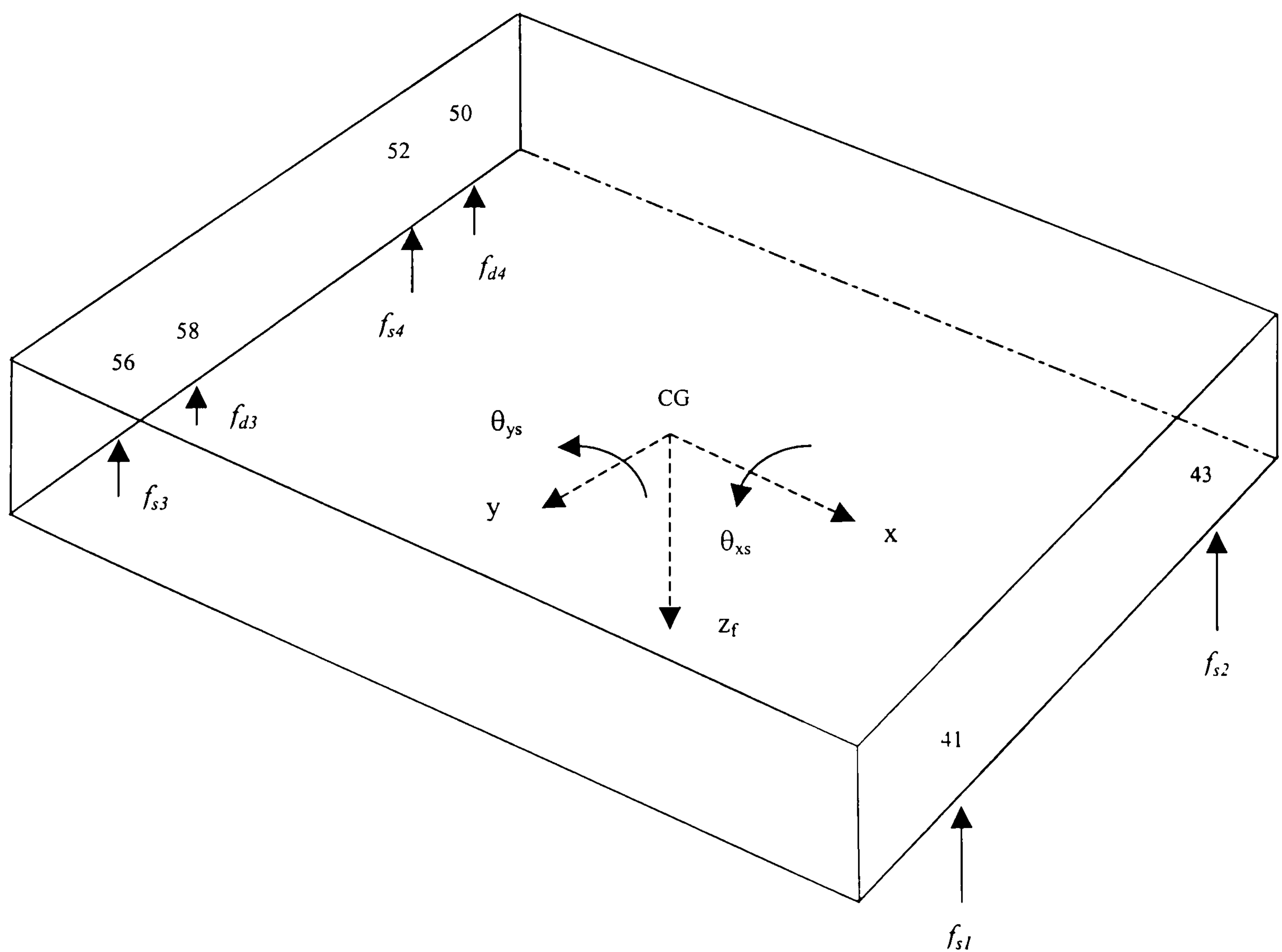


Fig. C3 A free-body diagram for the sprung mass

In bounce motion;

$$m_s \ddot{z}_s = - (f_{s1} + f_{s2} + f_{s3} + f_{d3} + f_{s4} + f_{d4}) + m_s g \quad (C1)$$

In roll motion;

$$I_{sxx} \ddot{\theta}_{xs} = - y_{n41} f_{s1} - y_{n43} f_{s2} - y_{n56} f_{s3} - y_{n58} f_{d3} - y_{n52} f_{s4} - y_{n50} f_{d4} \quad (C2)$$

In pitch motion;

$$I_{syy} \ddot{\theta}_{ys} = x_{n41} f_{s1} + x_{n43} f_{s2} + x_{n56} f_{s3} + x_{n58} f_{d3} + x_{n52} f_{s4} + x_{n50} f_{d4} \quad (C3)$$

Considering a free-body diagram for the front unsprung mass in Fig. C 4;

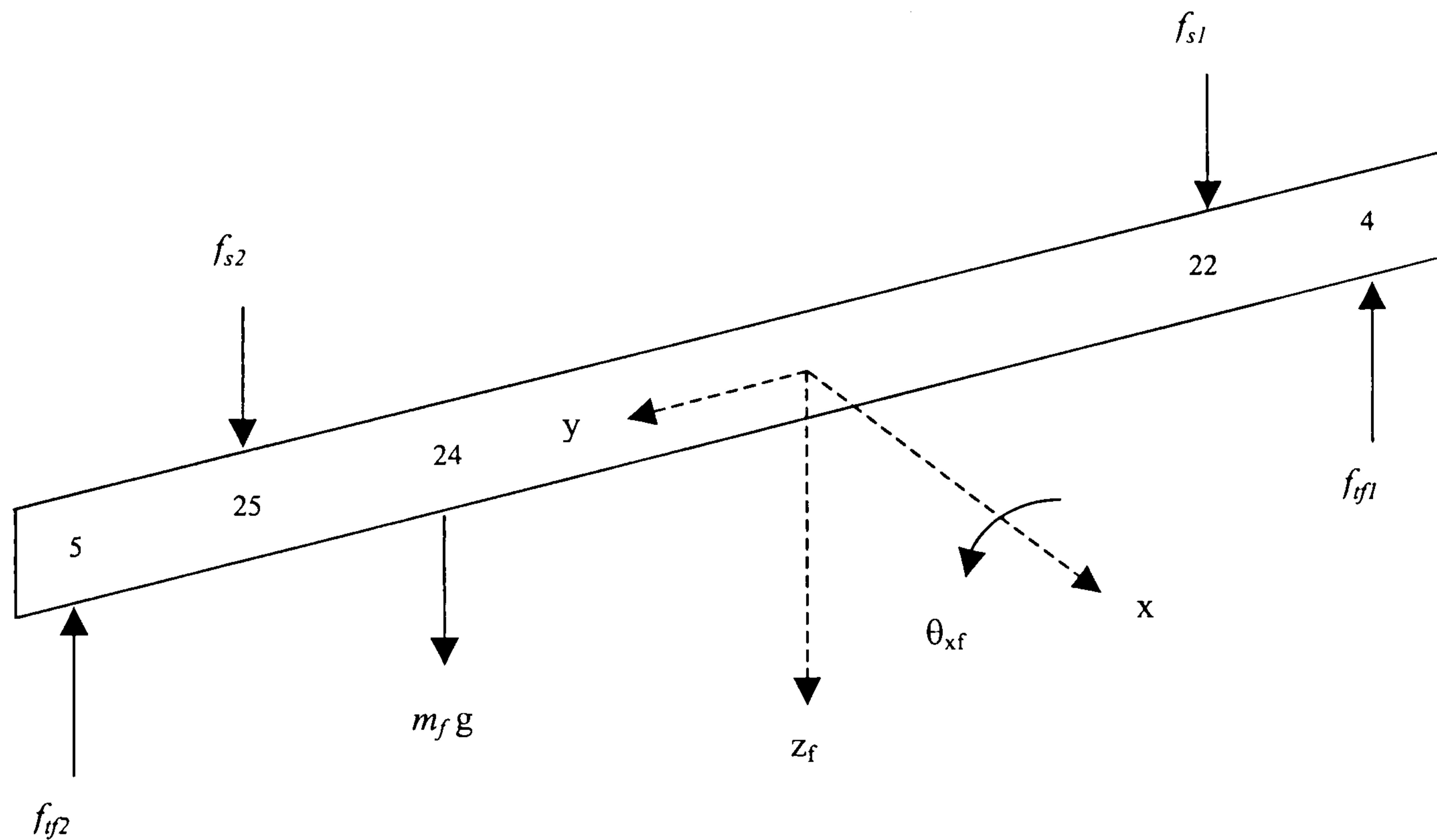


Fig. C4 A free-body diagram for the front axle

In bounce motion;

$$m_f \ddot{z}_f = - f_{if1} - f_{if2} + f_{s1} + f_{s2} + m_f g \quad (C4)$$

In roll motion;

$$I_{fxx} \ddot{\theta}_{xf} = - y_{n4} f_{if1} - y_{n5} f_{if2} + y_{n22} f_{s1} + y_{n25} f_{s2} + y_{n24} m_f g \quad (C5)$$

Considering a free-body diagram for the rear unsprung mass in Fig. C5;

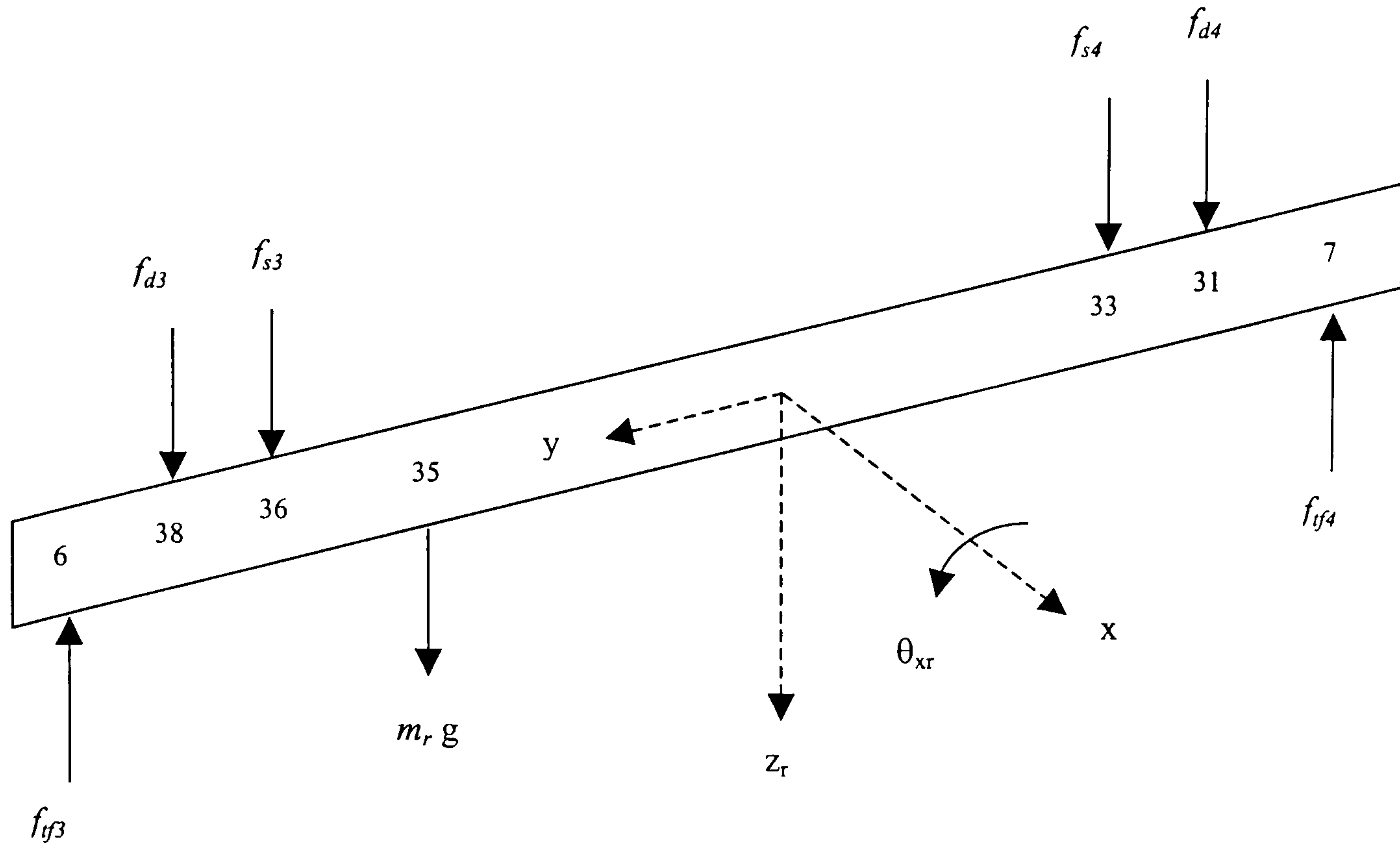


Fig. C5 A free-body diagram for the rear axle

In bounce motion;

$$m_r \ddot{z}_r = f_{d3} + f_{s3} + f_{d4} + f_{s4} + m_r g - f_{tr3} - f_{tr4} \quad (C6)$$

In roll motion;

$$I_{rxx} \ddot{\theta}_{xr} = -f_{tr3} y_{n6} - f_{tr4} y_{n7} + m_r g y_{n35} + f_{s3} y_{n36} + f_{d3} y_{n38} + f_{s4} y_{n33} + f_{d4} y_{n31} \quad (C7)$$

where

$$f_{if1} = -k_{if1}(z_{gf1} - (z_f + y_{n4}\theta_{xf})) \quad (C8)$$

$$f_{if2} = -k_{if2}(z_{gf2} - (z_f + y_{n5}\theta_{xf})) \quad (C9)$$

$$f_{if3} = -k_{if3}(z_{gr3} - (z_r + y_{n6}\theta_{xr})) \quad (C10)$$

$$f_{tr4} = -k_{tr4}(z_{gr4} - (z_r + y_{n7}\theta_{xr})) \quad (C11)$$

$$\begin{aligned}
f_{s1} = & -k_{sf1}((z_f + y_{n22}\theta_{xf}) - (z_s + y_{n41}\theta_{xs} - x_{n41}\theta_{ys})) \\
& -d_{f1}((\dot{z}_f + y_{n22}\dot{\theta}_{xf}) - (\dot{z}_s + y_{n41}\dot{\theta}_{xs} - x_{n41}\dot{\theta}_{ys}))
\end{aligned} \tag{C12}$$

$$\begin{aligned}
f_{s2} = & -k_{sf2}((z_f + y_{n25}\theta_{xf}) - (z_s + y_{n43}\theta_{xs} - x_{n43}\theta_{ys})) \\
& -d_{f2}((\dot{z}_f + y_{n25}\dot{\theta}_{xf}) - (\dot{z}_s + y_{n43}\dot{\theta}_{xs} - x_{n43}\dot{\theta}_{ys}))
\end{aligned} \tag{C13}$$

$$f_{s3} = -k_{sr3}((z_r + y_{n36}\theta_{xr}) - (z_s + y_{n56}\theta_{xs} - x_{n56}\theta_{ys})) \tag{C14}$$

$$f_{d3} = -d_{r3}((\dot{z}_r + y_{n38}\dot{\theta}_{xr}) - (\dot{z}_s + y_{n58}\dot{\theta}_{xs} - x_{n58}\dot{\theta}_{ys})) \tag{C15}$$

$$f_{s4} = -k_{sr4}((z_r + y_{n33}\theta_{xr}) - (z_s + y_{n52}\theta_{xs} - x_{n52}\theta_{ys})) \tag{C16}$$

$$f_{d4} = -d_{r4}((\dot{z}_r + y_{n31}\dot{\theta}_{xr}) - (\dot{z}_s + y_{n50}\dot{\theta}_{xs} - x_{n50}\dot{\theta}_{ys})) \tag{C17}$$

C3. Estimation Results

Model 2

The results of fitting Model 2 to the experimental data are presented in Fig. C6. The results are similar to that of Model 1. The model predicted the vehicle motion relatively well for both body bounce and wheel hop modes. However, similarly to Model 1, it did not take into account the resonance of the engine. The differences in the amplitude ratios and phases are again mainly due to model simplicity and the optimised set of model parameters that were obtained according to the cost function formulated in the optimisation process.

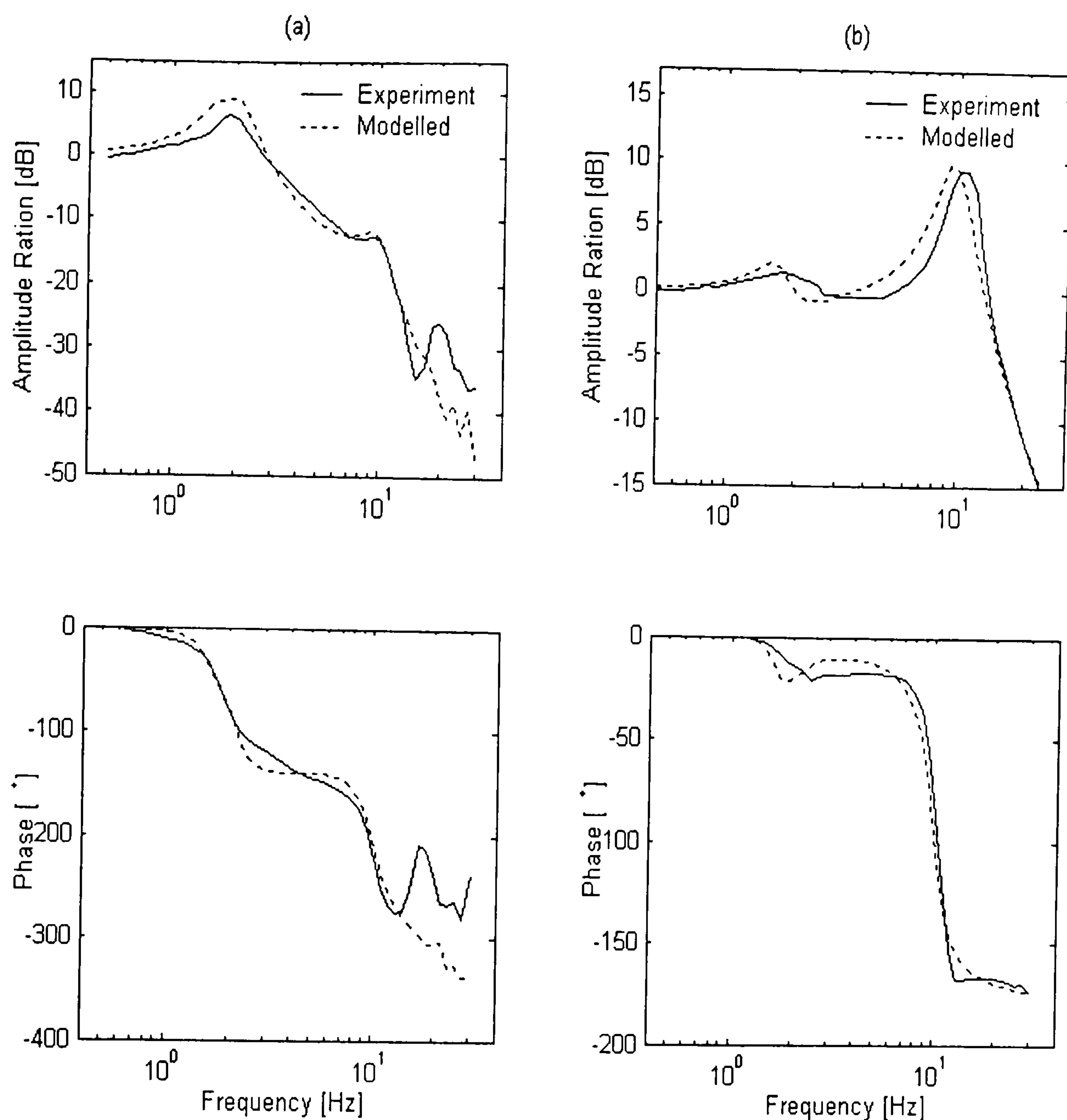


Fig. C6 Fitting result of vehicle frequency response by Model 2; (a) vehicle body, (b) front axle.

Model 3

The results of fitting Model 3 to the data are presented in Fig. C7 and Fig. C8. Part (a) of Fig. C7 shows a comparison of the vehicle body frequency response at the vehicle rear end (the vehicle towing-point), and part (b) of the figure at the front end (front chassis cross member). From the figure, it can be said that the model captured the vehicle body motion at the rear and front ends relatively well, with, again, greater errors toward high frequencies, and the front end showed the greatest difference.

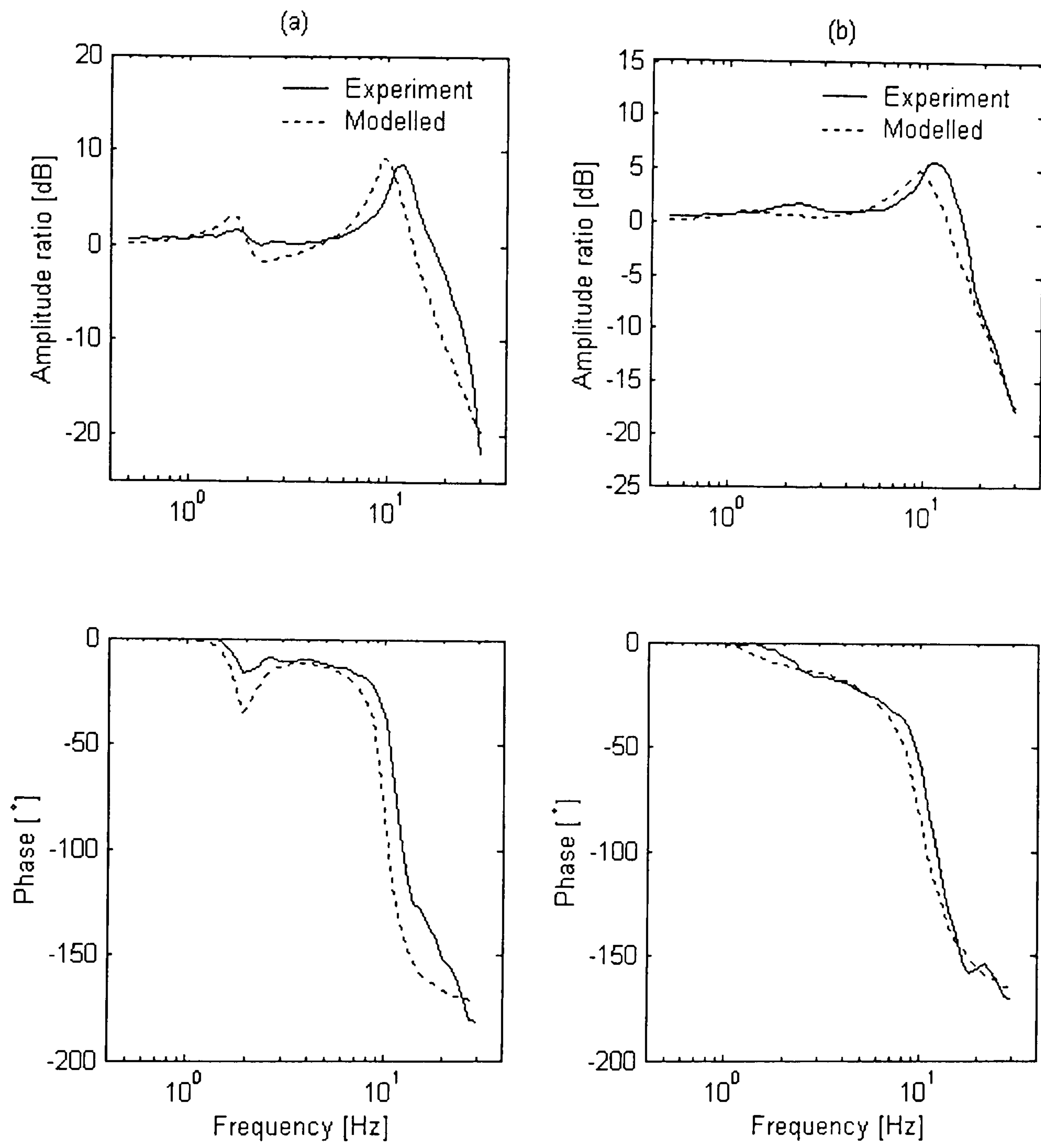


Fig. C7 Fitting result of vehicle body frequency response by Model 3; (a) rear end, (b) front end.

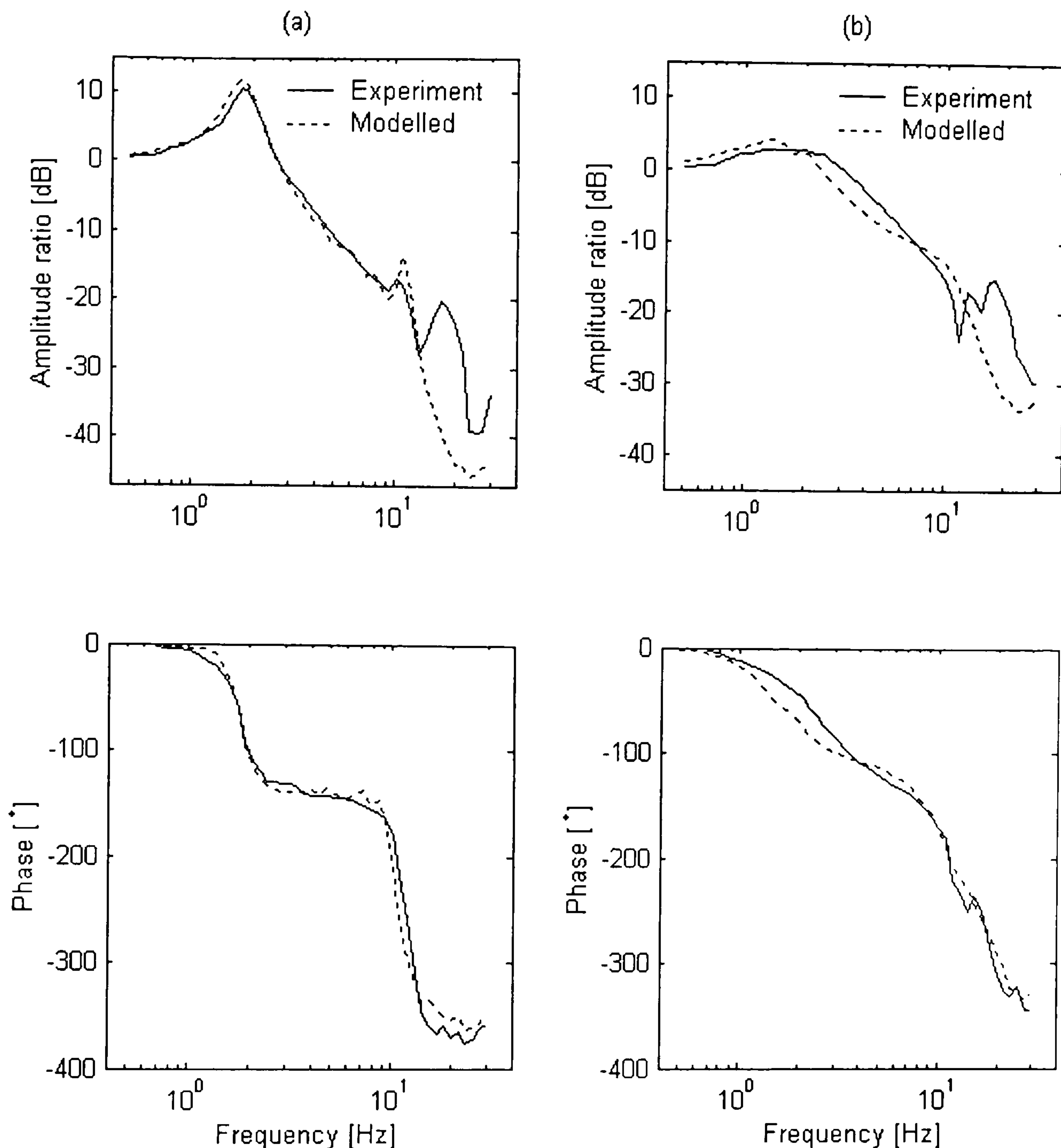


Fig. C8 Fitting result of vehicle frequency response by Model 3; (a) rear axle, (b) front axle.

A comparison of the modelled and experimentally obtained frequency responses of the vehicle rear axle and front axle are presented respectively in part (a) and (b) of Fig. C8. From the plots, the differences between the two by comparing the curves can be clearly seen, it can however be said that the model was able to capture the dynamics of the vehicle rear and front axles. Additionally to the errors due to the

model assumptions, the main difference between the experimental and modelled ones were due to the estimated model parameters obtained, which were possibly biased to some of the frequency points when formulating the cost function in the parameter estimation procedure.

C4. Discussion of results

Comparing the results of fitting the model frequency responses to the vehicle experimental frequency responses by Model 1, Model 2 and Model 3, some comments can be made. First, Model 2, which was a planar model that attempted to take into account the effect of the suspension linkages, however it was not easy and straightforward to tell the difference between the quality of fit of Model 2 and Model 1 from the plots in Fig. 5.5.1 and Fig. C6. Model 1 benefited from being simpler and can be derived by hand and required less computational time and effort when solving the equations of motion to obtain the model responses. The model required less specific information about the vehicle, so that Model 1 can be used more widely with different types of vehicle. Model 1 also contained fewer parameters to be identified, hence required less computational time in the parameter estimation process. Model 3, a 7DOF model which took into account the vehicle body roll, pitch and bounce, and the axles bounce and roll, contained greater number of parameters to be estimated, thus was more time consuming when compared to that of Model 1 and Model 2. Each model has its advantages and disadvantages over the other two. To choose one from these three depends on the purpose of model, and this should reflect the best compromise between the modelling effort and time consumed to complete the identification and parameter estimation task.